Large Space-Time Scale Behavior of Linearly Interacting Diffusions

een wetenschappelijke proeve op het gebied
van de Natuurwetenschappen, Wiskunde
en Informatica

Proefschrift

ter verkrijging van de graad van doctor aan de Katholieke Universiteit Nijmegen, volgens besluit van het College van Decanen in het openbaar te verdedigen op maandag 8 november 1999, des namiddags om 3.30 uur precies door

Jan Meinderts Swart,

geboren op 7 juni 1970 te Drachten.

Promotor: Prof. Dr. W.Th.F. den Hollander

Manuscriptcommissie: Prof. Dr. A. Greven (Universität Erlangen-Nürnberg)
Prof. Dr. J. Gärtner (Technische Universität Berlin)
Dr. J.D.M. Maassen

This work is part of the research programme of the 'Stichting voor Fundamenteel Onderzoek der Materie' (FOM) and the 'Nederlandse Organisatie voor Wetenschappelijk Onderzoek' (NWO).

ISBN

Large Space-Time Scale Behavior of Linearly Interacting Diffusions

doctoral thesis

Jan M. Swart

November 1999

Chapter 2 of this dissertation has appeared as 'F. den Hollander and J.M. Swart, Renormalization of hierarchically interacting isotropic diffusions. J. Stat. Phys. 93 (1998) 243–291.'

Chapter 3 of this dissertation has appeared as an internal report of the Mathematical Institute of the University of Nijmegen at http://www-math.sci.kun.nl/math/onderzoek/reports/reports_nl.html.

Preface

In the world around us we often find systems consisting of a large number of identical components. For example, one may think of a material that is made out of molecules, a population that consist of individuals, or a tissue that is made out of cells. Typically, the behavior of the individual components in such a system can be described by a few relatively easy laws, and these laws in principle determine the behavior of the whole system. Nevertheless, questions about the collective behavior of the system as a whole may be highly non-trivial to answer on the basis of the local laws. It is such questions that the discipline of interacting particle systems occupies itself with.

A typical feature in the description of such systems is the use of probability theory. Our information on the world around us is never complete, and probability theory tells us how to describe this incompleteness. For systems with only few interacting components, it often happens that we can idealize our incomplete information by complete information, but for systems with many interacting components this is never the case.

Instead, it is customary to idealize such systems in a different way. Instead of studying large finite systems, one often studies infinite systems, the philosophy being that many particles sometimes behave almost as infinitely many particles. The phenomena of interest in the infinite system often occur in large but finite systems only with a high probability, on appropriate time scales, or on approximate values of certain parameters.

Mathematicians mostly prefer to work with systems in which the local laws are of an extremely simple nature. Such 'toy models' all occur as extremely simplified caricatures of physical systems. Examples are the Ising model (describing a ferromagnet), independent bond percolation (describing the transport of a fluid through a porous medium), the voter model (describing voting behavior, or the distribution of genetic types), the contact process (describing an epedemic) and the exclusion process (describing a gas). Despite their apparent simplicity, these models turn out to give better descriptions of the systems they are supposed to model than one might think at first sight. The reason is that the collective behavior of large systems is often to a large extent insensitive to details of the local laws. Thus, it often happens that the laws governing the behavior of systems on large scales are universal in a large class of models, defined by different local laws. Understanding this universality is some sort of a holy grail in the field.

This dissertation is about understanding one particular form of universality that occurs in a certain class of models. The models that we occupy ourselves with consist of an infinite system of diffusion processes, interacting through a linear drift. They share some properties with the voter model, such as a phase transition between so-called 'stable' and 'clustering' behavior, the importance of random walk representations, and the key role played by dual models.

Because of the use of (rather heavy) diffusion theory, the description of these diffusion models is a bit more involved than that of most other 'toy models'. Once this work is done, however, one is rewarded for the extra effort by the fact that the models immediately allow for a number of generalizations. These generalizations make it possible to investigate the already mentioned phenonemon of universality. For some models we are able to prove that the laws governing their behavior on large space and time scales are indeed universal in certain classes of local laws.

This dissertation is the collection of three research papers. Each paper can be read independently of the others. There is a certain overlap, even a slight inconsistency in notation between the papers, but I hope the reader will accept these in return for having a –necessarily temporary– view on how this problem area and our views on it evolved.

A dissertation is meant to address a diverse audience, of whom some are experts in the field, others friends who are just curious to know what I have been doing during the last four years. For those who have a certain general knowledge on probability and analysis (say, who are not afraid of the words: Markov process, σ-field, separable metric space), but who have no or only a limited experience with diffusion processes, I have included an appendix explaining the basic ideas of the theory. I strongly recommend any reader who does not know what a local drift and diffusion function are to at least take a look there (page 155) before trying to read anything else. I have also included an introductory chapter, where I can give a little more attention to the origin of the diffusion models (as limits of discrete particle models) than the restricted space of a scientific article allows. There I have also included some notes on the sometimes strange and curved ways by which we discovered certain results, and on how our views on the subject changed along the way.

Contents

Chapter 1

Introduction

1.1 Particle models and diffusion limits

In this section we will see how the interacting diffusion models that are the subject of this dissertation arise as the continuum limits of certain discrete models. We begin by introducing a simple model, consisting of an urn containing balls of different colors, whose contents are subject to a certain random time evolution. Next, we extend this model step by step, until we arrive at a system of infinitely many interacting urns, which –in the limit that each urn contains many balls– is described by an interacting diffusion model of the type that we study in this dissertation.

1.1.1 The p-type q-tuple model

Consider an urn containing N balls of p colors. We introduce the following *resampling mechanism*. Each set of q balls in the urn (a 'q-tuple') is selected with rate $q(q-1)$. We take them out of the urn and look at their colors. If two or more of the q balls are of the same color, then we simply place all q balls back in the urn. But if all the q balls have different colors, we throw one of these balls (randomly selected) away, and replace it by a (random) ball with the same color as one of the remaining $q-1$ balls. The new set of q balls that we get in this way we place back in the urn. For every q-tuple of balls we repeat this procedure ad infinitum.

We are interested in the frequencies of the p different colors in our urn as time evolves. For $\alpha = 1, \ldots, p$, let us write $Y_\alpha(t)$ for the number of balls of color α at time t. Thus, we consider a stochastic process

$$Y = (Y(t))_{t\geq 0} = (Y_\alpha(t))_{t\geq 0,\ \alpha=1,\ldots,p} \tag{1.1.1}$$

taking values in the space

$$S_N := \Big\{x \in \mathbb{N}^p : \sum_\alpha x_\alpha = N\Big\}. \tag{1.1.2}$$

The process Y is a *jump process*. We denote its transition probabilities by

$$P_t(x, y) := P[Y(s+t) = y \mid Y(s) = x] \tag{1.1.3}$$

and we write $(P_t)_{t\geq 0}$ for the semigroup, acting on real functions on S_N, given by

$$(P_t f)(x) := \sum_{y \in S_N} P_t(x, y) f(y). \tag{1.1.4}$$

The generator G of $(P_t)_{t\geq 0}$ is given by

$$(Gf)(x) = \sum_{\substack{\Delta\subset\{1,\ldots,p\} \\ |\Delta|=q}} \Big(\prod_{\gamma\in\Delta} x_\gamma\Big) \sum_{\substack{\alpha,\beta\in\Delta \\ \alpha\neq\beta}} \Big(f(x + e_\alpha - e_\beta) - f(x)\Big). \tag{1.1.5}$$

Here the first sum ranges over all subsets Δ of $\{1, \ldots, p\}$ that contain exactly q elements, and we have defined

$$e_\alpha = (0, 0, \ldots, 0, 1, 0, \ldots, 0), \tag{1.1.6}$$

with the 1 in the α-th position. If x describes the composition of our urn at a given moment, then $\prod_{\gamma\in\Delta} x_\gamma$ is the number of q-tuples in the urn that carry the colors in Δ, and $\frac{1}{q(q-1)}$ is the probability that in a given q-tuple a ball of color α is replaced by a ball of color β. Since each q-tuple is selected with rate $q(q-1)$, the latter factor cancels, and in our urn balls of color α are replaced by balls of color β with rate

$$\sum_{\substack{\Delta\subset\{1,\ldots,p\} \\ |\Delta|=q \\ \Delta\ni\alpha,\beta}} \Big(\prod_{\gamma\in\Delta} x_\gamma\Big). \tag{1.1.7}$$

This explains formula (1.1.5).

1.1.2 The diffusion limit

We are interested in the behavior of the process Y for large N. In order to get a nontrivial limit, we must rescale our process. We look at the process $X^N = (X^N_\alpha(t))_{t\geq 0,\ \alpha=1,\ldots,p}$, defined as

$$X^N_\alpha(t) := \frac{1}{N} Y_\alpha(N^{2-q} t). \tag{1.1.8}$$

The process $X^N(t)$ takes values in the discrete $(p-1)$-dimensional simplex

$$K_p^N := \Big\{ x \in \tfrac{1}{N}\mathbb{N}^p : \sum_\alpha x_\alpha = 1 \Big\}. \tag{1.1.9}$$

We can write the generator of X^N as

$$(G^N f)(x) = \sum_{y \in K_p^N} \Gamma^N(x,y) f(y), \tag{1.1.10}$$

where

$$\Gamma^N(x,y) = N^2 \sum_{\substack{\Delta \subset \{1,\dots,p\} \\ |\Delta| = q}} \Big(\prod_{\gamma \in \Delta} x_\gamma \Big) \sum_{\substack{\alpha,\beta \in \Delta \\ \alpha \neq \beta}} \Big(\delta(x + \tfrac{1}{N} e_\alpha - \tfrac{1}{N} e_\beta, y) - \delta(x,y) \Big). \tag{1.1.11}$$

As N tends to infinity, we expect the process X^N to converge to a *diffusion process* X on the $(p-1)$-dimensional simplex

$$K_p := \Big\{ x \in [0,\infty)^p \ : \ \sum_\alpha x_\alpha = 1 \Big\}. \tag{1.1.12}$$

In order to find out what the generator of X could be, we have to calculate the moments of the kernel Γ^N in (1.1.11) up to leading order in N.

It is immediately clear from the definition that the zeroth moment of Γ^N is zero:

$$\sum_{y \in K_p^N} \Gamma^N(x,y) = 0. \tag{1.1.13}$$

For the first moment, we note that the process X^N is a martingale: in our resampling procedure the expected increase in the number of balls of any color is zero. This implies that for all $\alpha = 1, \dots, p$

$$\sum_{y \in K_p^N} \Gamma^N(x,y)(y_\alpha - x_\alpha) = 0. \tag{1.1.14}$$

For the second moments a small calculation yields

$$\begin{aligned} &\sum_{y \in K_p^N} (y_\alpha - x_\alpha)(y_\beta - x_\beta) \sum_{\substack{\gamma,\eta \in \Delta \\ \gamma \neq \eta}} \delta(x + \tfrac{1}{N} e_\gamma - \tfrac{1}{N} e_\eta, y) \\ &\quad = \sum_{\substack{\gamma,\eta \in \Delta \\ \gamma \neq \eta}} \tfrac{1}{N^2} (\delta_{\alpha\gamma}\delta_{\beta\gamma} + \delta_{\alpha\eta}\delta_{\beta\eta} - \delta_{\alpha,\gamma}\delta_{\beta,\eta} - \delta_{\alpha,\eta}\delta_{\beta,\gamma}) \\ &\quad = \frac{2}{N^2} 1_{\{\alpha,\beta \in \Delta\}} (q\delta_{\alpha\beta} - 1), \end{aligned} \tag{1.1.15}$$

so that for all $\alpha, \beta = 1, \ldots, p$

$$\sum_{y \in K_p^N} \Gamma^N(x, y)(y_\alpha - x_\alpha)(y_\beta - x_\beta) = \sum_{\substack{\Delta \subset \{1,\ldots,p\} \\ |\Delta| = q \\ \Delta \ni \alpha,\beta}} \Big(\prod_{\gamma \in \Delta} x_\gamma\Big) 2\Big(q\delta_{\alpha\beta} - 1\Big). \tag{1.1.16}$$

Finally, we have

$$\sum_{y \in K_p^N} \Gamma^N(x, y)|y - x|^3 = \mathcal{O}(N^{-1}), \tag{1.1.17}$$

uniformly in x as $N \to \infty$. Using a Taylor expansion one can now check[1] that for all $f \in \mathcal{C}^2(\mathbb{R}^p)$

$$\lim_{N \to \infty} \sup_{x \in K_p^N} \Big|(G^N f)(x) - (A_{p,q} f)(x)\Big| = 0, \tag{1.1.18}$$

where we have defined

$$(A_{p,q} f)(x) := \sum_{\substack{\Delta \subset \{1,\ldots,p\} \\ |\Delta| = q}} \Big(\prod_{\gamma \in \Delta} x_\gamma\Big) \sum_{\alpha,\beta \in \Delta} \Big(q\delta_{\alpha\beta} - 1\Big) \tfrac{\partial^2}{\partial x_\alpha \partial x_\beta} f(x) \tag{1.1.19}$$

By definition, the domain of $A_{p,q}$ is

$$\mathcal{D}(A_{p,q}) := \mathcal{C}^2(K_p), \tag{1.1.20}$$

the space of real functions on K_p that can be extended to a function in $\mathcal{C}^2(\mathbb{R}^p)$. Formula (1.1.18) makes clear that the definition of $A_{p,q} f$ does not depend on the choice of this extension.

We conclude from (1.1.18) that if the jump process X^N converges to a diffusion process X, then the generator of X has to be an extension of $A_{p,q}$. It would carry too far for this introduction to prove the convergence of X^N to X, but we take (1.1.18) as our motivation to study solutions to the martingale problem for $A_{p,q}$. It can be shown that this martingale probem is well-posed.[2]

---

[1] Compare [16], Theorem 1.1 of chapter 10.

[2] Uniqueness of solutions to the martingale problem for $A_{p,q}$ can be shown with the help of techniques mentioned in the proof of Example 3.1.8 below. Convergence of X^N to X (in the sense of weak convergence on path space $\mathcal{D}_{K_p}[0, \infty)$) is a non-trivial problem, even when uniqueness of solutions to the martingale problem for $A_{p,q}$ is known. The problem is to show tightness for X^N. It is sufficient if the closure of $A_{p,q}$ generates a Feller semigroup, see [16], section 8 of chapter 4. It is known that the closure of $A_{p,2}$ generates a Feller semigroup, see [16], Theorem 2.8 of chapter 8.

Because of the restriction $\sum_\alpha x_\alpha = 1$, the system of coordinates $(x_\alpha)_{\alpha=1,\ldots,p}$ is overdetermined. The first $p-1$ coordinates suffice, and we may identify K_p with the space

$$\hat{K}_p := \Big\{x \in [0,\infty)^{p-1} \;:\; \sum_{\alpha=1}^{p-1} x_\alpha \leq 1\Big\}. \tag{1.1.21}$$

A function $\hat{f} \in \mathcal{C}^2(\hat{K}_p)$ we can extend to a function $f \in \mathcal{C}^2(\mathbb{R}^p)$ in such a way that

$$f(x_1,\ldots,x_p) = \hat{f}(x_1,\ldots,x_{p-1}) \quad \forall (x_1,\ldots,x_{p-1}) \in \hat{K}_p,\ x_p \in \mathbb{R}. \tag{1.1.22}$$

This function f has the property that

$$\begin{aligned} &\tfrac{\partial}{\partial x_\alpha} f(x_1,\ldots,x_p) = \tfrac{\partial}{\partial x_\alpha} \hat{f}(x_1,\ldots,x_{p-1}) \qquad (\alpha = 1,\ldots,p-1) \\ &\tfrac{\partial}{\partial x_p} f(x_1,\ldots,x_p) = 0, \end{aligned} \tag{1.1.23}$$

and hence we see that in terms of the restricted system of coordinates $(x_\alpha)_{\alpha=1,\ldots,p-1}$ the operator $A_{p,q}$ must be written as

$$(A_{p,q}\hat{f})(x) = \sum_{\substack{\Delta \subset \{1,\ldots,p\} \\ |\Delta| = q}} g_\Delta(x) \sum_{\alpha,\beta \in \Delta\backslash\{p\}} \Big(q\delta_{\alpha\beta} - 1\Big) \tfrac{\partial^2}{\partial x_\alpha \partial x_\beta} \hat{f}(x), \tag{1.1.24}$$

with

$$g_\Delta(x) = \begin{cases} \Big(\prod_{\gamma\in\Delta} x_\gamma\Big) & \text{if } p \notin \Delta \\ \Big(\prod_{\gamma\in\Delta\backslash\{p\}} x_\gamma\Big)\Big(1 - \sum_\gamma x_\gamma\Big) & \text{if } p \in \Delta. \end{cases} \tag{1.1.25}$$

The system of coordinates $(x_\alpha)_{\alpha=1,\ldots p-1}$ has the advantage that it is not overdetermined, but since it violates the symmetry between the colors, the formula for $A_{p,q}$ is more complicated.

1.1.3 Examples

We consider two examples of p-type q-tuple diffusion models in more detail. The first example is the p-type 2-tuple model. For this model formula (1.1.19) can be written as

$$
\begin{aligned}
(A_{p,2}f)(x) &= \sum_{\substack{\Delta\subset\{1,\dots,p\}\\ |\Delta|=2}} \Big(\prod_{\gamma\in\Delta} x_\gamma\Big) \sum_{\alpha,\beta\in\Delta} \Big(2\delta_{\alpha\beta}-1\Big)\tfrac{\partial^2}{\partial x_\alpha \partial x_\beta} f(x) \\
&= \sum_{\alpha,\beta\in\Delta} \sum_{\substack{\Delta\subset\{1,\dots,p\}\\ |\Delta|=2\\ \Delta\ni\alpha,\beta}} \Big(\prod_{\gamma\in\Delta} x_\gamma\Big)\Big(2\delta_{\alpha\beta}-1\Big)\tfrac{\partial^2}{\partial x_\alpha \partial x_\beta} f(x).
\end{aligned}
\tag{1.1.26}
$$

For $\alpha \neq \beta$, the only set Δ occuring in the second summation is $\Delta = \{\alpha, \beta\}$ and the summand simplifies to

$$
-x_\alpha x_\beta \tfrac{\partial^2}{\partial x_\alpha \partial x_\beta} f(x). \tag{1.1.27}
$$

For $\alpha = \beta$, the summand in (1.1.26) can be written as

$$
\sum_{\gamma\neq\alpha} x_\alpha x_\gamma (2-1) \tfrac{\partial^2}{\partial {x_\alpha}^2} f(x) = x_\alpha(1-x_\alpha)\tfrac{\partial^2}{\partial {x_\alpha}^2} f(x). \tag{1.1.28}
$$

Thus, we can write $A_{p,2}$ as

$$
(A_{p,2}f)(x) = \sum_{\alpha\beta} x_\alpha(\delta_{\alpha\beta} - x_\beta)\tfrac{\partial^2}{\partial x_\alpha \partial x_\beta} f(x). \tag{1.1.29}
$$

It is easy to see that $A_{p,2}$ takes the same form in the restricted system of coordinates $(x_\alpha)_{\alpha=1,\dots,p-1}$, only with the summation restricted to $\alpha, \beta \neq p$.

The function $w : \hat{K}_p \to \mathbb{R}^{p-1} \otimes \mathbb{R}^{p-1}$ given by

$$
w_{\alpha\beta}(x) := x_\alpha(\delta_{\alpha\beta} - x_\beta) \qquad (x \in \hat{K}_p) \tag{1.1.30}
$$

is called the Wright-Fisher diffusion matrix, and the diffusion process X associated to $A_{p,2}$ is (an example of) a *Wright-Fisher diffusion process*. The behavior of this process is well-known. If we start it in a point where all colors are present, then after a finite time one of the colors becomes extinct. It is clear from our resampling mechanism that once a color has become extinct it cannot reappear again, and therefore the process $X(t) = (X_\alpha(t))_{\alpha=1,\dots,p}$ moves from that moment on in the subspace

$$
F_\alpha := \{x \in K_p : x_\alpha = 0\}, \tag{1.1.31}
$$

where α is the color that has become extinct. F_α is called the α-th *face* of the simplex K_p. Note that F_α is isomorphic to K_{p-1}. After again a finite time a second color becomes extinct, and then another one, and the process moves in subspaces of ever lower dimension until only one color is left. At that moment the process gets stuck in one of the extremal points of the simplex. Thus, there is a finite stopping time τ such that almost surely

$$X(\tau+t)=X(\tau)\in\{e_1,\ldots,e_p\}\qquad\forall t\geq 0. \tag{1.1.32}$$

As a second example, we consider the p-type p-tuple model. For this model, the only set Δ occuring in the summation in (1.1.19) is $\Delta=\{1,\ldots,p\}$ and the formula simplifies to

$$(A_{p,p}f)(x)=\Big(\prod_{\gamma=1}^{p}x_\gamma\Big)\sum_{\alpha\beta}\Big(p\delta_{\alpha\beta}-1\Big)\tfrac{\partial^2}{\partial x_\alpha\partial x_\beta}f(x). \tag{1.1.33}$$

The differential operator occuring here looks more transparant when expressed in different coordinates. We choose an orthonormal basis $(e'_1,\ldots,e'_p)$ for $\mathbb{R}^p$ such that

$$e'_p=\tfrac{1}{\sqrt{p}}(1,1,\ldots,1), \tag{1.1.34}$$

and we write $x'_1,\ldots x'_p$ for the coordinates of a point x in this new basis:

$$x=\sum_\alpha x_\alpha e_\alpha=\sum_\alpha x'_\alpha e'_\alpha\qquad(x\in\mathbb{R}^p). \tag{1.1.35}$$

Then

$$\tfrac{\partial}{\partial x'_p}f(x)=\lim_{\varepsilon\to 0}\varepsilon^{-1}\Big(f(x+\varepsilon e'_p)-f(x)\Big)=\tfrac{1}{\sqrt{p}}\sum_\alpha\tfrac{\partial}{\partial x_\alpha}f(x) \tag{1.1.36}$$

and

$$\tfrac{\partial^2}{\partial {x'_p}^2}f(x)=\tfrac{1}{p}\Big(\sum_\alpha\tfrac{\partial}{\partial x_\alpha}\Big)^2f(x)=\tfrac{1}{p}\sum_{\alpha\beta}\tfrac{\partial^2}{\partial x_\alpha\partial x_\beta}f(x). \tag{1.1.37}$$

We note that the Laplacian Δ takes the same form in any orthonormal coordinate system:

$$(\Delta f)(x)=\sum_\alpha\tfrac{\partial^2}{\partial {x_\alpha}^2}f(x)=\sum_\alpha\tfrac{\partial^2}{\partial {x'_\alpha}^2}f(x). \tag{1.1.38}$$

Combining (1.1.37) and (1.1.38) we see that we can rewrite (1.1.33) as

$$(A_{p,p}f)(x)=p\Big(\prod_{\gamma=1}^{p}x_\gamma\Big)\sum_{\alpha=1}^{p-1}\tfrac{\partial^2}{\partial {x'_p}^2}f(x). \tag{1.1.39}$$

Here $\sum_{\alpha=1}^{p-1}\frac{\partial^2}{\partial {x'_p}^2}$ is the Laplacian in the plane given by the equation $\sum_\alpha x_\alpha=1$. Thus, $A_{p,p}$ is of the form

$$A_{p,p}=\sum_{\alpha,\beta=1}^{p-1}w_{\alpha\beta}(x)\tfrac{\partial^2}{\partial x'_\alpha\partial x'_\beta}\qquad\text{with}\qquad w_{\alpha\beta}(x)=\delta_{\alpha\beta}g(x), \tag{1.1.40}$$

and $g : K_p \to [0, \infty)$ some function. We express this by saying that the diffusion matrix w is *isotropic*. The process X associated to $A_{p,p}$ is an *isotropic* diffusion. Such an isotropic diffusion (with zero drift) is just a time-transformed Brownian motion. The behavior of X is as follows. In a finite time one of the types gets extinct and, as is clear from our resampling mechanism, after that time the frequencies of all other colors remain fixed. Thus, there is a finite stopping time τ such that almost surely

$$X(\tau + t) = X(\tau) \in \partial K_p \qquad \forall t \geq 0, \tag{1.1.41}$$

where

$$\partial K_p := \bigcup_{\alpha=1}^{p} F_\alpha = \{x \in K_p : x_\alpha = 0 \text{ for some } \alpha\}. \tag{1.1.42}$$

The behavior of general p-type q-tuple diffusions with $2 < q < p$ is similar to that of the two examples above. One by one colors become extinct, until only $q - 1$ colors are left and the process comes to a halt.

1.1.4 The p-type q-tuple model with migration

Once again consider an urn with balls of p colors, but this time let us assume that the number of balls is not fixed. Instead, we introduce the following *migration mechanism*. We assume that with rate $c\rho$ our urn receives balls from some large reservoir, where the proportions of the different colors are fixed to $(\theta_\alpha)_{\alpha=1,\ldots,p} = \theta \in K_p$. Moreover, we assume that each ball in our urn dissapears from the urn with rate c. Here $c, \rho \in (0, \infty)$ and it is easy to see that the expected number of balls in our urn tends to ρ as time tends to infinity.

In addition to this migration mechanism, we assume that the q-tuples present in our urn are at any given moment subject to the resampling mechanism described in section 1.1.1 with rate $q(q-1)\rho^{2-q}$. We write $Y^\rho_\alpha(t)$ for the number of balls of color α present in our urn at time t, and we define

$$X^\rho_\alpha(t) := \frac{1}{\rho} Y^\rho_\alpha(t). \tag{1.1.43}$$

The generator of the process X^ρ is given by (compare (1.1.8) and (1.1.11))

$$\begin{aligned}(G^\rho f)(x) = \rho^2 \sum_{\substack{\Delta\subset\{1,\ldots,p\}\\ |\Delta|=q}} \Big(\prod_{\gamma\in\Delta} x_\gamma\Big) \sum_{\substack{\alpha,\beta\in\Delta\\ \alpha\neq\beta}} \Big(f(x + \tfrac{1}{\rho}e_\alpha - \tfrac{1}{\rho}e_\beta) - f(x)\Big) \\ + \sum_\alpha \Big\{c\rho\theta_\alpha\Big(f(x + \tfrac{1}{\rho}e_\alpha) - f(x)\Big) + cx_\alpha\Big(f(x - \tfrac{1}{\rho}e_\alpha) - f(x)\Big)\Big\}.\end{aligned} \tag{1.1.44}$$

Here the first term results from the resampling mechanism (see formula (1.1.5)). One can check that for every $f \in \mathcal{C}^2([0,\infty)^p)$ and for every compact subset $C \subset [0,\infty)^p$

$$\lim_{\rho\to\infty} \sup_{x\in C\cap\frac{1}{\rho}\mathbb{N}^p} \Big|(G^\rho f)(x) - (A_{p,q}f)(x) - (B_{\theta,c}f)(x)\Big| = 0, \tag{1.1.45}$$

where $A_{p,q}$ is defined as in (1.1.19) and

$$(B_{\theta,c}f)(x) := c\sum_\alpha (\theta_\alpha - x_\alpha)\tfrac{\partial}{\partial x_\alpha} f(x). \tag{1.1.46}$$

Thus we expect the process X^ρ to converge, as $\rho \to \infty$, to a process X with generator G, where G is an extension of the operator $A_{p,q} + B_{\theta,c}$. It is possible to prove existence of solutions to the martingale problem for $A_{p,q} + B_{\theta,c}$, and one can check that for a solution X with initial condition $X(0) \in K_p$

$$X(t) \in K_p \qquad \forall t \geq 0 \tag{1.1.47}$$

almost surely. This means that in the limit of large ρ, the total number of balls in the urn is approximately fixed to ρ. For a function $f \in \mathcal{C}^2(K_p)$ we define $A_{p,q}f + B_{\theta,c}f$ by extending f to a function in $\mathcal{C}^2([0,\infty)^p)$, and we can see that the result does not depend on the choice of the extension.

1.1.5 Uniqueness problems

Remarkably, it is not known whether the diffusion processes introduced in the last section are well-defined. Namely, it is not known in general whether solutions to the martingale problem for $A_{p,q}+B_{\theta,c}$ are unique. The standard result about strong uniqueness of solutions to stochastic differential equations does not apply, because it is not possible to find a Lipschitz continuous root of the diffusion matrix occuring in $A_{p,q}$ (see section B.2). The problems occur at the boundary ∂K_p of the simplex. It is possible to represent solutions to the martingale problem for $A_{p,q} + B_{\theta,c}$ as solutions to a stochastic differential equation of the form

$$dX_\alpha(t) = c(\theta_\alpha - X_\alpha(t))dt + \sum_\beta \sigma_{\alpha\beta}(X(t))dB_\beta(t), \tag{1.1.48}$$

where the function σ is locally Lipschitz on $K_p\backslash\partial K_p$ but not Lipschitz at ∂K_p. Therefore strong uniqueness of solutions to (1.1.48) can be shown only up to the first hitting time of ∂K_p.

If $c = 0$, then solutions to (1.1.48) are martingales, and hence after the first hitting time of ∂K_p the process stays in one of the faces F_α. Since these faces are isomorfic to K_{p-1}, it is then possible to prove strong uniqueness of solutions to (1.1.48) by induction. (For this technique, see Example 3.1.8 below.)

For c sufficiently large, it is possible to show that solutions to the martingale problem for $A_{p,q} + B_{\theta,c}$ never hit the boundary ∂K_p, so that solutions to (1.1.48) are strongly unique for all time. (For this technique, see Theorem 2.2.9 below.)

However, for c positive but not too large the process X with probability one reaches the boundary ∂K_p in a finite time. In fact, it hits ∂K infinitely often in a finite time, each time bouncing back from it. About processes with such behavior very little is known. Certainly the standard strong uniqueness results do not apply, except in the one-dimensional (i.e. 2-type) case.

For the p-type 2-tuple model, there are several ways to circumvent this problem. For example, it is possible to find a root σ that is lower-triangular. This corresponds the fact that each subselection of the colors is itself a Markov process (following a Wright-Fisher diffusion). Thus, the idea is that one can first prove strong uniqueness for one color, using one-dimensional techniques, then prove strong uniqueness for the second color conditional on the first one, and so on.

In another approach, one can prove weak uniqueness for the p-type 2-tuple model with migration by more or less explicitly calculating all moments of $X(t)$. Here one uses the fact that $A_{p,2} + B_{\theta,c}$ maps a polynomial of degree n into a polynomial of degree at most n. Thus, the time evolution of all moments up to n-th order is described by a closed system of equations, that is easily seen to have a unique solution.[3] This technique has the advantage that it also proves that the closure of $A_{p,2} + B_{\theta,c}$ generates a Feller semigroup.

For general p-type q-tuple models, one can see that $A_{p,q}$ maps a polynomial of degree n into a polynomial of degree $n - 2 + q$, and hence for $q \geq 3$ the time evolution of moments up to n-th order cannot be expressed in a closed system. It seems that duality techniques (involving moments) that are known to work in certain other models also fail here, and uniqueness of solutions to the martingale problem for $A_{p,q} + B_{\theta,c}$ with $p, q \geq 3$, for general $\theta \in K_p$ and $c > 0$, is still an open problem.

1.1.6 Interacting p-type q-tuple models

We now consider a collection of urns, indexed by a finite or countable Abelian group Λ, with

$$\begin{array}{rl} \text{group operation} & i + j \\ \text{inverse} & -i \\ \text{unit element} & 0. \end{array} \tag{1.1.49}$$

[3] In fact, this solution can be represented in terms of a dual process, which is sometimes handy in calculations. But here this duality is not essential.

For example, Λ may be the n-dimensional integer lattice $\mathbb{Z}^n$, or a finite part of $\mathbb{Z}^n$ with periodic boundary conditions. We will also frequently consider the case that $\Lambda = \Omega_N$, the N-dimensional hierarhical group (see below).

We fill the urns according to a Poisson process, where balls of color α occur with intensity $\rho\theta_\alpha$, with $(\theta_\alpha)_{\alpha=1,\ldots,p} = \theta \in K_p$ and $\rho > 0$. Thus the total number of balls in each urn is Poisson distributed with mean ρ, and a given ball is with probability θ_α of color α. We introduce the following migration mechanism between our urns. We assume that balls independently of each other perform continuous-time random walks on Λ, where a ball in urn $j \in \Lambda$ jumps to urn $i \in \Lambda$ with rate

$$a(j-i). \tag{1.1.50}$$

Here the *migration kernel* $a : \Lambda \to [0,\infty)$ is a function satisfying

$$\sum_i a(i) < \infty. \tag{1.1.51}$$

We further assume that every q-tuple of balls present at a certain moment in an urn is subject to the resampling mechanism descibed in section 1.1.1 with rate ρ^{2-q}. Let us write $Y_i^{\rho,\alpha}(t)$ for the number of balls of color α in urn i at time t, and let us consider the process

$$X^\rho = (X^\rho(t))_{t\geq 0} = (X_i^{\rho,\alpha}(t))_{i\in\Lambda,\ t\geq 0}^{\alpha=1,\ldots,p} \tag{1.1.52}$$

given by

$$X_i^{\rho,\alpha}(t) := \frac{1}{\rho} Y_i^\alpha(t). \tag{1.1.53}$$

Then we expect X^ρ to converge, as $\rho \to \infty$, to a diffusion process

$$X = (X(t))_{t\geq 0} = (X_i^\alpha(t))_{i\in\Lambda,\ t\geq 0}^{\alpha=1,\ldots,p}, \tag{1.1.54}$$

with initial condition

$$X_i(0) = \theta \qquad (i \in \Lambda), \tag{1.1.55}$$

that solves the martingale problem for the operator A, given by

$$\begin{aligned}(Af)(x) :=& \sum_i \sum_{\substack{\Delta\subset\{1,\ldots,p\}\\ |\Delta|=q}} \Big(\prod_{\gamma\in\Delta} x_\gamma\Big) \sum_{\alpha,\beta\in\Delta} \Big(q\delta_{\alpha\beta} - 1\Big) \tfrac{\partial^2}{\partial x_i^\alpha \partial x_i^\beta} f(x) \\ &+ \sum_{ij,\alpha} a(j-i)(x_j^\alpha - x_i^\alpha) \tfrac{\partial}{\partial x_i^\alpha} f(x).\end{aligned} \tag{1.1.56}$$

The domain of A is the space of all $\mathcal{C}^2$-functions on $(K_p)^\Lambda$ that depend on finitely many coordinates only, and a point $x \in (K_p)^\Lambda$ we denote as

$$x = (x_i^\alpha)_{i\in\Lambda}^{\alpha=1,\dots,p}. \tag{1.1.57}$$

Infinite systems of interacting diffusion processes of this type and their generalizations are the main subject of study in this dissertation.

For definiteness let us write down the operator A above in restricted coordinates. With $\hat{K}_p$ as in (1.1.21) and $(\hat{K}_p)^\Lambda$ the space of all points $x = (x_i)_{i\in\Lambda}$ with $x_i \in \hat{K}_p$, we have for any $\mathcal{C}^2$-function $\hat{f}$ that depends on finitely many of the x_i only

$$\begin{aligned}(A\hat{f})(x) :=& \sum_i \sum_{\alpha\beta} w_{\alpha\beta}^{(p,q)}(x_i) \tfrac{\partial^2}{\partial x_i^\alpha \partial x_i^\beta} f(x) \\ &+ \sum_{ij,\alpha} a(j-i)(x_j^\alpha - x_i^\alpha) \tfrac{\partial}{\partial x_i^\alpha} f(x),\end{aligned} \tag{1.1.58}$$

where for any $x \in \hat{K}_p$ and $\alpha, \beta = 1, \dots, p-1$

$$w_{\alpha\beta}^{(p,q)}(x) = \sum_{\substack{\Delta\subset\{1,\dots,p\} \\ |\Delta|=q \\ \Delta\ni\alpha,\beta}} g_\Delta(x)(q\delta_{\alpha\beta} - 1), \tag{1.1.59}$$

with $g_\Delta(x)$ as in (1.1.25). A short look at section 1.1.3 learns us that in particular

$$\begin{aligned} w_{\alpha\beta}^{(p,2)}(x) &= x_\alpha(\delta_{\alpha\beta} - x_\beta) \\ w_{\alpha\beta}^{(p,p)}(x) &= \Big(\prod_\gamma x_\gamma\Big)\Big(1 - \sum_\gamma x_\gamma\Big)(p\delta_{\alpha\beta} - 1). \end{aligned} \tag{1.1.60}$$

1.1.7 Other models

We briefly mention here some diffusion models that are closely related to the p-type q-tuple models. We start with two models with non-compact state space.

Feller's branching diffusion Consider an urn with balls of one color. With rate one each ball is with equal probabilities replaced by either two or zero balls ('critical branching'). In the right scaling, this process converges to a diffusion on $[0, \infty)$ whose generator extends the operator

$$(Af)(x) := x \tfrac{\partial^2}{\partial x^2} f(x). \tag{1.1.61}$$

Extensions to more colors are immediate. These models can be viewed as a limit of the p-type 2-tuple model when one of the types occupies almost all the urn.

It is easy to see that the evolution of the rare types can then be approximated by independent critical branching.

Mutually catalytic branching This model has been introduced by Dawson and Perkins [14]. Consider an urn with balls of two colors, subject to the following resampling mechanism. With rate one, each pair (2-tuple) of balls is selected. If they have different colors, then one ball is selected and with equal probability this ball is replaced by two or zero balls of the same color. In the right scaling, we expect this process to converge to a diffusion on $[0,\infty)^2$ whose generator extends the operator

$$(Af)(x) := x_1x_2\big(\tfrac{\partial^2}{\partial x_1{}^2}+\tfrac{\partial^2}{\partial x_2{}^2}\big)f(x). \tag{1.1.62}$$

This is an isotropic model. It can be viewed as a limit of the 3-type 3-tuple model when one of the types is much more common than the other two. The evolution of the two rare colors can then be approximated by mutually catalytic branching. It is known that the martingale problem for the operator A in (1.1.62) is well-posed, also when a migration term $B_{\theta,c}$ as in (1.1.46) is added to (1.1.62). In fact, it is easy to see that all the moments of the process can be calculated. However, since the moment problem is not well-posed in this non-compact setting, one has to do a bit more to prove uniqueness of solutions to the martingale problem. This is achieved in [14] by means of a self-duality of the model, due to Mytnik.

The 4-type $2+2$-tuple model Consider an urn with balls of four colors. With rate one each quadruple of balls is selected. If all the colors are different, then with equal probabilities either the ball with color 1 is replaced by a ball of color 2 or vice versa, or the ball of color 3 is replaced by a ball of color 4 or vice versa. In this way, the proportion of the colors 1 plus 2 with respect to the colors 3 plus 4 is not changed. We start in a situation where the total number of balls of the colors 1 plus 2 equals the total number of balls of the colors 3 plus 4, and we denote by $Y_\alpha(t)$ the number of balls of color α at time t. In the right scaling, we expect the process (Y_1, Y_3) to converge to a diffusion on $[0,1]^2$ whose generator extends the operator

$$(Af)(x) := x_1(1-x_1)x_3(1-x_3)\big(\tfrac{\partial^2}{\partial x_1{}^2}+\tfrac{\partial^2}{\partial x_3{}^2}\big)f(x). \tag{1.1.63}$$

This is an isotropic model, similar in spirit to mutually catalytic branching. In fact, in the limit that the colors 2 and 4 are much more common than the colors 1 and 3, we recover the mutually catalytic model. When a migration term $B_{\theta,c}$ as in (1.1.46) is added to the diffusion in (1.1.63), it is not known whether solutions to the corresponding martingale problem are always unique.

Composition-dependent resampling This is not really one model, but a recipe by which one can produce a whole collection of other models. One considers the situation where the rate with which the various resampling mechanisms take place depends on the whole composition of the urn. This includes all the models discussed so far. For example, if in the 3-type 2-tuple model one lets the rate at which one color replaces another depend (linearly) on the amount of the third color present, then one arrives at the 3-type 3-tuple model. If in the 2-type 2-tuple model one lets the resampling rate depend on the amount of both types present in the urn, then one finds the operator

$$(Af)(x) := x^2(1-x)^2\frac{\partial^2}{\partial x^2}, \tag{1.1.64}$$

known as Kimura's random selection model. *One of the main goals in this dissertation is to show that for infinite systems of interacting diffusions such modifications of the diffusion function do not influence the behavior of the system on large scales, both in space and time.*

1.2 Overview of the three articles

1.2.1 Renormalization theory

Renormalization theory is one of the most succesful techniques for understanding universal large scale behavior of interacting particle systems, at least on the level of heuristic and non-rigorous calculations. The basic idea of the theory is quite simple. First, one needs to find a way to describe a system on a series of ever larger scales. A usual way is to group the particles into blocks, consisting of a particle and a few of its neighbours, then group these blocks into larger blocks, and so on. With each scale is associated a set of variables describing the system as if viewed from ever larger distances, where details of the local behavior become ever less visible. For example, the first set of variables may give the precise state of each particle, the second set only the average value of all particles in a block, and the third set only averages over blocks of blocks, etc. Each time one goes to a larger scale, the probability law describing the new variables is a marginal of the law describing the old variables. Thus, in principle, one has a map describing how to go from the old variables to the new larger scale variables. This map is called a *renormalization transformation.* If it is the case that under iteration of this transformation different local laws converge to one and the same global law, then one has universal behavior on large scales.

In practice it is not so easy to realize this renormalization scheme. In order to make it work, one needs an efficient way to describe the probability law of the renormalized variables. However, it often happens that while the law of the local variables has nice properties, the renormalized law has not (for example, it may be non-Markovian or non-Gibbsian). In such cases a rigorous study of the renormalization transformation is very hard and frequently impossible.

In some special cases we are lucky and the renormalized system admits a nice description. Apart from the fact that it is nice that at least sometimes rigorous renormalization calculations are possible, the study of these cases is also interesting from a more fundamental point of view. If we understand better why universality occurs here, we may also find ways to understand systems for which the renormalization scheme does not work so nicely.

The research contained in this dissertation started in 1995, inspired by just such an example of a system for which the renormalization scheme works. This is a system of linearly interacting diffusions on the hierarchical group, introduced in the next section.

1.2.2 Renormalization of interacting diffusions

By definition, the N-dimensional hierarchical group is

$$\Omega_N := \Big\{(i_k)_{k=1,2,\ldots} : i_k \in \{0,\ldots,N-1\},\ i_k \neq 0 \text{ finitely often}\Big\}. \tag{1.2.1}$$

With componentwise addition modulo N, this is a countable Abelian group. We denote the origin by $0 = (0,0,\ldots)$. Think of $i \in \Omega_N$ as an address: then i_1 is the house number, i_2 the street, i_3 the town, and so on. Ω_N is ordered in a hierarchical way, where N houses form a street, N streets form a town, N towns form a province, and so on. One defines

$$\|i\| := \min\{k \in \mathbb{N} : i_l = 0 \ \ \forall l > k\}. \tag{1.2.2}$$

We call $\|i - j\|$ the hierarchical distance between i and j. For example, if i and j are in the same town, but not in the same street, they are at hierarchical distance 2.

Now let us imagine that each $i \in \Omega_N$ represents an urn with balls of p colors, and let us write $X_i^\alpha(t)$ for the relative frequency of color α in urn i at time t. We consider the process

$$X^N = (X^N(t))_{t\geq 0} = (X_i^{N,\alpha}(t))_{t\geq 0,\ i\in\Omega_N}^{\alpha=1,\ldots,p-1}, \tag{1.2.3}$$

where $X_i^N(t)$ takes values in $\hat{K}_p$. For reasons that will become clear later we choose to denote the N-dependence of our process explicitly. We assume that the initial frequencies of the colors are described by some $\theta \in \hat{K}_p$

$$X_i^N(0) = \theta \qquad (i \in \Omega_N). \tag{1.2.4}$$

If the urns are subject to a resampling mechanism and a migration mechanism as described in section 1.1.6, and the total number of balls in each urn is large, then we expect X^N to solve the martingale problem for an operator of the form

$$(Af)(x) := \sum_{ij} a(j-i) \sum_{\alpha} (x_j^\alpha - x_i^\alpha) \tfrac{\partial}{\partial x_i^\alpha} f(x) + \sum_i \sum_{\alpha\beta} w_{\alpha\beta}(x_i) \tfrac{\partial^2}{\partial x_i^\alpha \partial x_i^\beta} f(x), \tag{1.2.5}$$

with domain the $\mathcal{C}^2$-functions f that depend on finitely many x_i^α only. Here the diffusion matrix w can be the p-type q-tuple diffusion matrix $w^{(p,q)}$, originating from a q-tuple resampling mechanism, but we also allow for more general w, originating from a composition-dependent resampling mechanism.

We choose the migration kernel a in such a way that the strength of the migration between two urns depends only on their hierarchical distance. The collection of all urns at hierarchical distance at most k from an urn i

$$\{j \in \Omega_N : \|j - i\| \leq k\} \tag{1.2.6}$$

we call the *k-block around i*. We fix constants $c_1, c_2, \ldots \in (0, \infty)$ and for all $k = 1, 2, \ldots$ we let the balls in our urns be subject to the following migration mechanism: With rate c_k/N^{k-1} each ball in an urn i chooses a random urn in the k-block around i (possibly itself) and migrates to that urn. This means that the migration kernel a is given by

$$a(i) = \sum_{k=\|i\|}^{\infty} \frac{c_k}{N^{2k-1}}. \tag{1.2.7}$$

To understand why (1.2.7) is the correct formula, note that a ball in urn i decides with rate c_k/N^{k-1} to jump to another urn in the k-block around i. If $k \geq \|i\|$, this urn is with probability N^{-k} the origin.

The process X^N can be represented, on an appropriately chosen probability space equipped with $(p-1)$-dimensional independent Brownian motions $(B_i)_{i\in\Omega_N}$, as a solution to the following system of stochastic differential equations:

$$dX_i^{N,\alpha}(t) = \sum_{k=1}^{\infty} \frac{c_k}{N^{k-1}} \Big(X_i^{N,k,\alpha}(t) - X_i^{N,\alpha}(t) \Big) dt + \sum_{\beta} \sigma_{\alpha\beta}(X_i^N(t)) dB_i^\beta(t)$$
$$(i \in \Omega_N,\ \alpha = 1, \ldots, d,\ t \geq 0), \tag{1.2.8}$$

where

$$\tfrac{1}{2} \sum_{\gamma} \sigma_{\alpha\gamma}(x) \sigma_{\beta\gamma}(x) = w_{\alpha\beta}(x) \tag{1.2.9}$$

and $X_i^{N,k}(t)$ is the *k-block average around i*:

$$X_i^{N,k,\alpha}(t) := N^{-k} \sum_{j:\|j-i\|\leq k} X_i^{N,\alpha}(t). \tag{1.2.10}$$

It is clear that the hierarchical group with its structure of blocks made out of smaller blocks is ideally suited for renormalization theory. For certain 2-type models it has been shown that the system in (1.2.8) admits a rigorous description in terms of a renormalization transformation, in the limit where the dimension N of the hierarchical group tends to infinity. Since we expect this result to hold more generally, we formulate it here as a non-rigorous conjecture.

For $c \in (0, \infty)$, $x \in \hat{K}_p$ and for any diffusion matrix w on $\hat{K}_p$, let us write $A_x^{w,c}$ for the operator

$$(A_x^{w,c} f)(y) := \sum_{\alpha} c(x_\alpha - y_\alpha) \tfrac{\partial}{\partial y_\alpha} f(y) + \sum_{\alpha\beta} w_{\alpha\beta}(y) \tfrac{\partial^2}{\partial y_\alpha \partial y_\beta} f(y). \tag{1.2.11}$$

We expect that for 'reasonable' w (this is one point where we are non-rigorous) the martingale problem for $A_x^{w,c}$ is well-posed and the associated diffusion process has a unique equilibrium and is ergodic. By $Z_x^{w,c}$ we denote the solution to the martingale problem for $A_x^{w,c}$ with initial condition $Z_x^{w,c}(0) = x$, and by $\nu_x^{w,c}(dy)$ we denote the equilibrium distribution associated with $A_x^{w,c}$.

For each $c \in (0, \infty)$ we define a renormalization transformation F_c, acting on diffusion matrices w (we are vague as to the precise domain of F_c) by the formula

$$(F_c w)_{\alpha\beta}(x) := \int_{\hat{K}_p} w_{\alpha\beta}(y) \nu_x^{w,c}(dy). \tag{1.2.12}$$

Conjecture 1.2.1 *Assume that X^N solves (1.2.8) with initial condition $X_i(0) = \theta$ for all $i \in \Omega_N$. Then for each $k \geq 0$*

$$(X_0^{N,k}(N^k t))_{t \geq 0} \Rightarrow (Z_\theta^{F^{(k)}w, c_{k+1}}(t))_{t \geq 0} \qquad \text{as } N \to \infty, \tag{1.2.13}$$

where $F^{(k)}w$ is the k-th iterate of renormalization transformations F_c applied to w:

$$F^{(k)}w := (F_{c_k} \circ \cdots \circ F_{c_1})w. \tag{1.2.14}$$

Furthermore, for any $t > 0$

$$(X_0^{N,k}(N^k t), \ldots, X_0^{N,0}(N^k t)) \Rightarrow (Z^k, \ldots, Z^0) \qquad \text{as } N \to \infty, \tag{1.2.15}$$

where $(Z^k, \ldots, Z^0)$ is a Markov chain (in this order!) with transition probabilities

$$P[Z^{n-1} \in dy | Z^n = x] = \nu_x^{F^{(n-1)}w, c_n}(dy) \qquad (n = 1, \ldots, k). \tag{1.2.16}$$

Note that in order to get a non-trivial limit in (1.2.13), we need to rescale space *and* time. While we rescale space by going to k-block variables, we rescale time by a factor N^k. In the limit $N \to \infty$ the block averages of differently sized blocks evolve on separate time scales. The average of a large block changes much slower than the average of a smaller block.

If we consider the time evolution of a k-block average, then we may treat its interaction (due to migration) with the much larger $(k+1)$-block that it is part of as if this $(k+1)$-block is an infinite reservoir in which the frequencies of colors are fixed. As we saw in section 1.1.4, the process $Z_x^{w,c}$ describes the behavior of an urn that is in interaction with such a reservoir.

After a sufficiently long time, the k-blocks reach equilibrium, subject to the value of the $(k+1)$-block with which they interact. The diffusion matrix describing the evolution of this $(k+1)$-block can then be found by averaging the diffusion matrix of the k-blocks with respect to this equilibrium distribution. This is how the renormalization transformation F_c arises. For the details behind the heuristics, we refer to Chapter 2.

1.2.3 A renormalization transformation

In [10], Dawson and Greven proved a rigorous form of Conjecture 1.2.1 for a class of 2-type models. They considered $X_i^N(t)$, indexed by the hierarchical group Ω_N and taking values in $\hat{K}_2 = [0,1]$, solving an equation of the form (compare (1.2.8))

$$dX_i^N(t) = \sum_{k=1}^{\infty} \frac{c_k}{N^{k-1}}\Big(X_i^{N,k}(t) - X_i^N(t)\Big)dt + \sqrt{g(X_i^N(t))}dB_i(t) \quad (t \geq 0,\ i \in \Omega_N), \tag{1.2.17}$$

where g is taken from the class $\mathcal{H}$ of functions $g : [0,1] \to [0,\infty)$ that are Lipschitz continuous and satisfy $g(x) = 0 \Leftrightarrow x \in \{0,1\}$. On $\mathcal{H}$ and for $c \in (0,\infty)$, the renormalization transformation $F_c : \mathcal{H} \to \mathcal{H}$ is defined by

$$(F_c g)(x) := \int_{[0,1]} g(y)\nu_x^{g,c}(dy), \tag{1.2.18}$$

with $\nu_x^{g,c}$ the unique equilibrium distribution of a diffusion whose generator extends

$$(A_x^{g,c} f)(y) := c(x-y)\tfrac{\partial}{\partial y} f(y) + g(y)\tfrac{\partial^2}{\partial y^2} f(y). \tag{1.2.19}$$

The equilibrium measure $\nu_x^{g,c}$ occuring in (1.2.18) is in fact known in closed form, and the transformation F_c is given by the following explicit formula:

$$(F_c g)(x) = \frac{\displaystyle\int_0^1 dy\, e^{-\int_x^y dz\, \frac{c(z-x)}{g(z)}}}{\displaystyle\int_0^1 dy\, \frac{1}{g(y)} e^{-\int_x^y dz\, \frac{c(z-x)}{g(z)}}}\ . \tag{1.2.20}$$

As we see from formula (1.2.20), $F_c g$ depends in a complicated and non-linear way on g, so it is not obvious how the iterates of F_c behave. In [1], Baillon, Clément, Greven and Den Hollander studied these iterates. They were able to show the following.

Proposition 1.2.2 *Let $g^* \in \mathcal{H}$ be given by*

$$g^*(x) := x(1-x). \tag{1.2.21}$$

Then g^* is a fixed shape under F_c, $c \in (0,\infty)$:

$$F_c(\lambda g^*) = \Big(\frac{c}{\lambda + c}\Big)\lambda g^* \qquad \forall \lambda > 0. \tag{1.2.22}$$

Moreover, for $k = 1, 2, \ldots$ assume that $c_k \in (0,\infty)$ satisfy

$$\sum_{k=1}^{\infty} \frac{1}{c_k} = \infty, \tag{1.2.23}$$

and define

$$\begin{aligned} \sigma_n &:= \sum_{k=1}^{n} \frac{1}{c_k} \\ F^{(n)}g &:= (F_{c_k} \circ \cdots \circ F_{c_1})g. \end{aligned} \tag{1.2.24}$$

Then for all $g \in \mathcal{H}$, one has

$$\lim_{n\to\infty} \sup_{x\in[0,1]} \Big| \sigma_n (F^{(n)}g)(x) - g^*(x) \Big| = 0. \tag{1.2.25}$$

Proposition 1.2.2 shows that the renormalization transformations F_c have a unique fixed shape g^* that attracts all diffusion functions $g \in \mathcal{H}$ after appropriate scaling. This implies that the system in (1.2.17) exhibits universal behavior on large space and time scales. As we already saw in section 1.2.2, the renormalized diffusion functions $F^{(k)}g$ describe the behavior of k-block averages on their appropriate time scales N^k. Thus, formula (1.2.25) shows that large block averages ($k \to \infty$) evolve according to the *universal* diffusion function g^*, independent of the diffusion function g of the individual components $X_i^N(t)$ of the system. The scale factors σ_n are not important here, because they can always be absorbed in a redefinition of the time scale. The same is true for the factor $c/(c+\lambda)$ in (1.2.22), so the fact that our renormalization transformation has a fixed shape instead of a fixed point is not important. It turns out that g^* is the Wright-Fisher diffusion function (see (1.1.30)), which arises in a natural way as the diffusion limit of the 2-type 2-tuple model.

Condition (1.2.23) is necessary for the universality observed in (1.2.25). It is known that condition (1.2.23) corresponds to *clustering* behavior in the model (Theorem 3 in [10]). This means that the components $X_i^N(t)$ of the system, after a long time, spend most of their time near the boundary of [0, 1]. In fact, according to Conjecture 1.2.1,

$$\lim_{N\to\infty} P[X_0^N(N^k t) \in dy | X_0^{N,k}(N^k t) = x] = K_x^{(k)}(dy), \tag{1.2.26}$$

where

$$K_x^{g,(k)}(dy) = \int\int\cdots\int \nu_x^{F^{(k-1)}g,c_k}(dz_1)\nu_{z_1}^{F^{(k-2)}g,c_{k-1}}(dz_2)\cdots\nu_{z_{k-1}}^{F^{(0)}g,c_1}(dy). \tag{1.2.27}$$

Thus, the probability measure $K_x^{g,(k)}(\cdot)$ describes the conditional distribution of the urns in a k-block, given that the k-block average is x. It has been shown in [1] that

$$K_\theta^{g,(k)}(\cdot) \Rightarrow (1-\theta)\delta_0 + \theta\delta_1 \qquad \text{as } k \to \infty, \tag{1.2.28}$$

if and only if (1.2.23) holds. Therefore (1.2.23) corresponds to the situation where after a long time each urn with probability close to θ contains almost only balls of color 1, and with probability close to $1-\theta$ almost only balls of color 2.

1.2.4 Higher-dimensional generalizations

The results in Proposition 1.2.2 leave one with a number of questions. Notably, one would like to understand better the origin of the observed universality. What is so special about the Wright-Fisher diffusion function that all other diffusion functions in the class $\mathcal{H}$ are attracted to it? In October 1995 my supervisor and me took this question as our motivation to study higher-dimensional equivalents of the transformation F_c.

For simplicity, we first restricted ourselves to isotropic diffusions. Thus, we considered a transformation of the form

$$(F_c g)(x) := \int_K g(y)\nu_x^{g,c}(dy), \tag{1.2.29}$$

where $K \subset \mathbb{R}^d$ is some compact and convex domain, and $\nu_x^{g,c}$ is the unique equilibrium distribution of a diffusion whose generator extends

$$(A_x^{g,c} f)(y) := \sum_\alpha c(x_\alpha - y_\alpha)\tfrac{\partial}{\partial y_\alpha} f(y) + g(y)\sum_\alpha \tfrac{\partial^2}{\partial {y_\alpha}^2} f(y). \tag{1.2.30}$$

This Ansatz immediately raised a number of questions.

1. From which class $\mathcal{H}$ can we choose our diffusion function g so that $F_c g$ is well-defined? In particular:
   (a) For which g is the martingale problem for the operator in (1.2.30) well-posed?
   (b) For which g does the associated diffusion have a unique equilibrium $\nu_x^{g,c}$?
   (c) Is it true that $F_c g \in \mathcal{H}$ for all $g \in \mathcal{H}$?
2. Do the iterates of the transformation F_c describe the large space-time scale behavior of a system of interacting diffusions, i.e., can we prove a rigorous version of Conjecture 1.2.1?
3. Does the transformation F_c have a unique fixed shape g^* that attracts all other $g \in \mathcal{H}$ after appropriate scaling?

In the search for answers to these questions, we were not completely without clues. First, of course, we knew that the Wright-Fisher diffusion arises in a special way as the diffusion limit of particle models (see section 1.1.2), and this might have something to do with its special role. Furthermore, we knew that condition (1.2.23), necessary for the universality, corresponds to clustering behavior. The fact that the components $X_i(t)$ of the system, after a long time, spend most of their time near the boundary of the domain K makes one suspect that the long-time behavior of the system does not 'feel' the diffusion function g on the interior of K, and from this the universality could possibly arise. In fact, one would suspect that an appropriate reasoning would possibly not even need to consider the iterations of a complicated transformation like the one in (1.2.20), but could perhaps understand the universal behavior by a direct reasoning based on the dynamics of the system.

We will see how much of this intuition turned out to be right... and how much wrong.

1.2.5 Renormalization of isotropic diffusions

We started our investigations by defining, in analogy with the one-dimensional case:

$$\mathcal{H} := \{g : K \to [0, \infty) \mid g \text{ Lipschitz}, g(x) = 0 \Leftrightarrow x \in \partial K\}, \tag{1.2.31}$$

where $\partial K := K \backslash K^\circ$ is the boundary of K, with K° the interior of K. Let us for the moment assume that questions 1 and 2 above can be solved, and let us first concentrate on question 3, which concerns the problem of understanding universality.

It is not clear from our definitions what function g^* could be a fixed shape under F_c, or in fact whether such a function exists at all. In the one-dimensional case, the fact that $g(x) = x(1-x)$ is a fixed shape follows from the equilibrium conditions for the time evolution of the first and second moments of solutions to the martingale problem for $A_x^{g,c}$. We found out that this proof immediately extends to the case that

$$\begin{aligned} K &= \{x \in \mathbb{R}^d : |x| \leq 1\} \\ g^*(x) &= 1 - |x|^2. \end{aligned} \tag{1.2.32}$$

In view of the interpretation of our model as explained in section 1.1, this case is rather unsatisfactory. We would rather like to be able to treat the p-type p-tuple model (see section 1.1.3) or the 4-type $2+2$-tuple model (see section 1.1.7). For the latter, $K = [0,1]^2$ and the diffusion function

$$g(x) = x_1(1-x_1)x_2(1-x_2) \tag{1.2.33}$$

arises in a natural way as the continuum limit of a discrete model. A natural idea would be to see if this function is a fixed shape under F_c. But this turns out not to be the case.

We found out that there is no explicit formula for F_c in dimensions $d \geq 2$. This has the following reason. The equilibrium $\nu_x^{g,c}$ solves the equation

$$\langle \nu_x^{g,c} | A_x^{g,c} f \rangle = 0 \qquad \forall f \in \mathcal{C}^2(K), \tag{1.2.34}$$

where we use the notation

$$\langle \mu | f \rangle := \int_K f(x)\mu(dx) \tag{1.2.35}$$

for any probability measure μ on K and any function $f \in \mathcal{C}(K)$. If $\nu_x^{g,c}$ has a sufficiently differentiable density, and also g is sufficiently smooth, then after an integration by parts we can rewrite (1.2.34) as

$$\langle (A_x^{g,c})^\dagger \nu_x^{g,c} | f \rangle = 0 \qquad \forall f \in \mathcal{C}^2(K), \tag{1.2.36}$$

so that for the density $\nu_x^{g,c}$ we find the partial differential equation

$$(A_x^{g,c})^\dagger \nu_x^{g,c} = \Big(-\sum_\alpha \tfrac{\partial}{\partial y_\alpha} c(x_\alpha - y_\alpha) + \sum_\alpha \tfrac{\partial^2}{\partial {y_\alpha}^2} g(y) \Big) \nu_x^{g,c}(y) = 0. \tag{1.2.37}$$

In vector notation we can write this equation in the form

$$\nabla \cdot (T_x^{g,c} \nu_x^{g,c}) = 0, \tag{1.2.38}$$

where $\cdot$ denotes inner product and

$$\begin{aligned} \nabla &:= (\tfrac{\partial}{\partial x_1}, \ldots, \tfrac{\partial}{\partial x_d}) \\ T &:= (T_1, \ldots, T_d) \\ (T_\alpha f)(y) &:= \Big(- c(x_\alpha - y_\alpha) + \tfrac{\partial}{\partial y_\alpha} g(y) \Big) f(y). \end{aligned} \tag{1.2.39}$$

The vector T can be interpreted as the expected flux, i.e., it measures the net transport of particles at each point in the domain. Equation (1.2.38) now says that *the divergence of the flux is zero.* In some cases, the equilibrium even turns out to solve the stronger equation

$$T_\alpha \nu_x^{g,c} = 0 \qquad\qquad (\alpha = 1, \ldots, d) \tag{1.2.40}$$

i.e., *the flux is zero.* In this case we say that the equilibrium is *reversible.* (This comes from the fact that the process, started in $\nu_x^{g,c}$, is symmetric with respect to time reversal.) A reversible equilibrium defines an L^2-space of square-integrable (with respect to $\nu_x^{g,c}$) functions, on which the operator $A_x^{g,c}$ is self-adjoint, which has many technical advantages. In the one-dimensional case (i.e., K is an interval), the equilibrium is always reversible, and equation (1.2.40) can easily be solved explicitly, leading to the explicit formula (1.2.20) for F_c. However, in the higher-dimensional case one can show that the equilibrium is for most choices of the parameters *not* reversible and all we have for $\nu_x^{g,c}$ is equation (1.2.34), which we do not know how to solve explicitly.

The idea that finally allowed us to get some control on F_c, at least on a heuristic level, was to try an expansion in c^{-1}. For large c, the equilibrium $\nu_x^{g,c}$ is sharply peaked around the point x. A simple moment calculation shows that for sufficiently differentiable g:

$$(F_c g)(x) = g(x) + c^{-1}\tfrac{1}{2} g(x)\Delta g(x) + \mathcal{O}(c^{-2}). \tag{1.2.41}$$

Here $\Delta := \sum_\alpha \frac{\partial^2}{\partial {x_\alpha}^2}$ is the Laplacian. If we want the right-hand side to be a multiple of g for all $c \in (0, \infty)$, then g has to be a solution of the Dirichlet equation

$$\Delta g(x) = \lambda \tag{1.2.42}$$

for some $\lambda \in \mathbb{R}$. With boundary conditions $g(x) = 0$ for $x \in \partial K$ this equation has a unique solution for each λ. Thus, we found out that the only possible candidate for a fixed shape are the multiples of the function $g^* \in \mathcal{C}^2(K^\circ) \cap \mathcal{C}(K)$, defined as the unique solution of

$$\begin{aligned} -\tfrac{1}{2}\Delta g^*(x) &= 1 \qquad \text{on } K^\circ \\ g^*(x) &= 0 \qquad \text{on } \partial K. \end{aligned} \tag{1.2.43}$$

In most cases (for example on $K = [0, 1]^2$) there is no explicit expression for g^*.

At this stage, one might think that maybe an abstract fixed-point theorem could be applied to establish on general grounds the existence of a fixed shape, and then (1.2.41) might perhaps be applied to identify this fixed shape as g^*. This is not the case. Such an approach would at best guarantee that the equation $F_c g = \lambda g$ has a solution for each $\lambda > 0$, but not that the solutions for different λ are scalar multiples of each other, which is what we want for a true fixed shape.

We finally resolved the issue because we found a way, although very heuristic in nature, to calculate higher-order terms in formula (1.2.41). This led to an expansion for $F_c g$ confirming that formula (1.2.22) carries over to the higher-dimensional situation, when g^* is as in (1.2.43). This heuristic reasoning was then replaced by a rigorous reasoning involving an operator $B^{-1}_{x,c}$ that is in some way an inverse to the operator

$$(B_{x,c}f)(y) := \sum_{\alpha} c(x_\alpha - y_\alpha)\tfrac{\partial}{\partial y_\alpha} f(y). \tag{1.2.44}$$

In fact, it turned out that the operator $B^{-1}_{x,c}$ was given by the formula

$$B^{-1}_{x,c} f = -\int_0^\infty T^{x,c}_t f\, dt, \tag{1.2.45}$$

where $(T^{x,c}_t)_{t\geq 0}$ is the Feller semigroup whose generator extends $B_{x,c}$. The essential fact that we were using in our proof of the fixed-shape property of g^* was that this semigroup carries functions of constant Laplacian over into functions of constant Laplacian.

Our techniques also implied that g^* is an attractive fixed shape. Thus, we could see that Proposition 1.2.2 generalizes completely to the higher-dimensional case. We also could prove convergence of the kernel $K^{g,(k)}_x(dy)$ (see (1.2.27)). In the higher-dimensional case, one has

$$K^{g,(k)}_x(\cdot) \Rightarrow K^{(\infty)}_x(\cdot) \qquad \text{as } k \to \infty \tag{1.2.46}$$

if and only if

$$\sum_{k=1}^{\infty} \frac{1}{c_k} = \infty, \tag{1.2.47}$$

where the limiting kernel $K^{(\infty)}_x(dy)$ is given by the distribution of Brownian motion, starting from x, when it first hits the boundary ∂K. This means that the system *clusters in a universal way:* the components $X^N_i(t)$ are, after a sufficiently long time, approximately distributed on ∂K according to probability distribution $K^{(\infty)}_\theta(\cdot)$, and this distribution is universal in all sequences c_k satisfying (3.1.5) and all $g \in \mathcal{H}$.

All these properties we were able to prove, assuming that we could answer the questions 1 (a)–(c) on page 21. Thus, in order to make our results complete we still needed to show those. This proved to be a real stumbling block.

For the choice of $\mathcal{H}$ in (1.2.31), we managed to solve question 1 (b): we could prove, provided that the martingale problem for $A_x^{g,c}$ is well-posed, that the associated diffusion has a unique equilibrium, solving (1.2.34), and is ergodic. The proof, completely different from the one that was known for the one-dimensional case, is functional analytic in nature and is based on rather heavy results from the theory of partial differential equations.

Questions 1 (a) and 1 (c) we were not able to solve. Concerning 1 (c), for example, we had the trouble that although we could prove that the equilibrium $\nu_x^{g,c}$ exists, is unique and solves (1.2.34), we had little control over it (for example no explicit formula). As a result we were only able to prove that $F_c g$ is a continuous function satisfying $F_c(x) = 0 \Leftrightarrow x \in \partial K$. We could not show Lipschitz continuity. The situation concerning question 1 (a) was even more disastrous. Because of the difficulties with uniqueness for higher-dimensional diffusions, explained in section 1.1.4, we were not able to show uniqueness of our diffusion for all $g \in \mathcal{H}$. We had this problem even for g^*, the only exception being the trivial case[4]

$$\begin{aligned} &K = \{x \in \mathbb{R}^d : |x| \leq 1\} \\ &g^*(x) = 1 - |x|^2. \end{aligned} \tag{1.2.48}$$

Summarizing we can say that we were able to answer question 3 completely, question 1 only partly, while question 2 stayed out of reach. Consequenctly, the results on question 3 remained somewhat up in the air.

After struggling with these problems for two years, we wrote down our partial results in the article 'Renormalization of Hierarchically Interacting Isotropic Diffusions' [21], which is Chapter 2 of this dissertation.

1.2.6 Non-isotropic models

In spite of the technical difficulties, our results had given us some insight into the origin of the large scale universality of our systems. Notably, we had found out the following.

1. The fixed shape is not always the 'natural' diffusion matrix that arises as the diffusion limit of discrete urn models.

2. It is important to look at the evolution of functions of constant Laplacian under the semigroup $(T_t^{x,c})_{t\geq 0}$ generated by the drift term in the opertor $A_x^{g,c}$.

[4] For uniqueness in this case, see formula (A.3.11) in the Appendix.

3. Apart from $F^{(k)}g$, it is also interesting to look at convergence of the kernel $K_x^{g,(k)}(dy)$ associated with clustering of the process.

With this information, we next tried to understand non-isotropic models. We knew that for general non-isotropic models there is no longer just one fixed shape. For example, on the 2-dimensional simplex

$$\hat{K}_3 := \Big\{x \in \mathbb{R}^2 : x_1, x_2 \geq 0, \quad x_1 + x_2 \leq 1\Big\} \tag{1.2.49}$$

one can check that the diffusion matrices

$$\begin{aligned} w^{(3,2)}_{\alpha\beta}(x) &:= x_\alpha(\delta_{\alpha\beta} - x_\beta) \\ w^{(3,3)}_{\alpha\beta}(x) &:= x_1x_2(1 - x_1 - x_2)(3\delta_{\alpha\beta} - 1), \end{aligned} \tag{1.2.50}$$

corresponding to the 3-type 2-tuple and the 3-type 3-tuple model, respectively (see section 1.1.3), are both fixed shapes under the transformation F_c defined in (1.2.12). For $w^{(3,2)}$ this follows from a simple moment calculation, while for $w^{(3,3)}$ it follows from our theory for isotropic diffusions, and the fact that

$$-\tfrac{1}{2} \sum_{\alpha,\beta=1}^{2} (3\delta_{\alpha\beta} - 1)\tfrac{\partial^2}{\partial x_\alpha \partial x_\beta}\, x_1x_2(1 - x_1 - x_2) = 1. \tag{1.2.51}$$

Here the differential operator $\sum_{\alpha,\beta=1}^{2}(3\delta_{\alpha\beta} - 1)\frac{\partial^2}{\partial x_\alpha \partial x_\beta}$ plays the same role as the Laplacian in (1.2.43); in fact it can be transformed into the Laplacian by a simple change of coordinates (see section 1.1.3).

We managed to mould the reasonings for $w^{(3,2)}$ and $w^{(3,3)}$ into a unified form. The key objects to look at turned out to be the so-called w-harmonic functions. If w is some given diffusion matrix on a domain K, then we say that a function $f \in \mathcal{C}^2(K)$ is w-harmonic if

$$\sum_{\alpha\beta} w_{\alpha\beta}(x)\tfrac{\partial^2}{\partial x_\alpha \partial x_\beta} f(x) = 0. \tag{1.2.52}$$

It turns out that the only $w^{(p,2)}$-harmonic functions are the affine functions $x \mapsto \sum_\alpha c_\alpha x_\alpha + c$, with c_α, c constants. For an isotropic diffusion matrix w, the w-harmonic functions are all functions f of zero Laplacian: $\Delta f = 0$; these are what are normally called harmonic functions. In both cases, the w-harmonic functions have the property that

$$f \text{ is } w\text{-harmonic} \Rightarrow \; T_t^{x,c} f \text{ is } w\text{-harmonic}, \tag{1.2.53}$$

where

$$(T_t^{x,c} f)(y) = f(x + (y - x)e^{-ct}) \tag{1.2.54}$$

is the semigroup generated by the operator $c\sum_\alpha (x_\alpha - y_\alpha)\frac{\partial}{\partial y_\alpha}$, i.e., the drift term in $A^{w,c}_x$. If (1.2.53) holds we say that a diffusion matrix w has *invariant harmonics*. This condition guarantees that systems of linearly interacting diffusions with given w-harmonic functions cluster in a universal way, as we explain now.

1.2.7 Harmonic functions and clustering

Let $X = (X_i(t))_{t\geq 0,\ i\in\Lambda}$ be a family of stochastic processes, indexed by an arbitrary Abelian group Λ, solving the martingale problem for an operator A of the form

$$(Af)(x) := \sum_{ij,\alpha} a(j-i)(x^\alpha_j - x^\alpha_i)\frac{\partial}{\partial x^\alpha_i} f(x) + \sum_i \sum_{\alpha\beta} w_{\alpha\beta}(x_i)\frac{\partial^2}{\partial x^\alpha_i \partial x^\beta_i} f(x). \quad (1.2.55)$$

If X has initial condition

$$X_i(0) = \theta \qquad (i \in \Lambda), \quad (1.2.56)$$

and if the diffusion matrix w has invariant harmonics, then it turns out that

$$E[f(X_i(t))] = f(\theta) \qquad \forall t \geq 0,\ i \in \Lambda,\ f\ w\text{-harmonic}. \quad (1.2.57)$$

If $w = w^{(p,2)}$, then the only w-harmonic functions are the affine functions, and (1.2.57) says nothing more than that the mean of $X_i(t)$ is conserved. In general, we express (1.2.57) by saying that the *harmonic mean* of $X_i(t)$ is conserved.

It turns out that the long-time behavior of the process $X = (X(t))_{t\geq 0} = (X_i(t))_{t\geq 0,\ i\in\Lambda}$ depends on the migration kernel a. If the *symmetrized random walk* on Λ, i.e., the random walk that jumps from i to j with rate

$$a(j-i) + a(i-j) \quad (1.2.58)$$

is *recurrent* and irreversible, then we believe that

$$X(t) \Rightarrow X(\infty) \qquad \text{as } t \to \infty, \quad (1.2.59)$$

where $X(\infty)$ has the following properties

$$\begin{aligned} &P[X_i(\infty) \in \partial_w K \ \ \forall i] = 1 \\ &P[X_i(t) = X_j(t) \ \ \forall i,j] = 1. \end{aligned} \quad (1.2.60)$$

Here

$$\partial_w K := \{x \in K : w(x) = 0\} \quad (1.2.61)$$

is what we call the *effective boundary* of the domain K. For example, for a p-type q-tuple model, $\partial_w \hat{K}_p$ consists of those compositions of the urn in which less than q

colors are present, so that the resampling process has come to a halt. On the other hand, if the symmetrized random walk on Λ is *transient,* then one can prove that (1.2.60) cannot hold.

For the p-type 2-tuple model we can understand the importance of the symmetrized random walk as follows. Consider two balls, drawn at a certain time t at random from two different urns or from one and the same urn. Both of these balls have in the past migrated according to the random walk with kernel a, and have been subject to the resampling mechanism described in section 1.1.1. If they have been introduced into the urn as a result of the resampling mechnism, then we call the ball whose color they copied their parent. The probability that our two balls are of the same color now depends on the probability that they descend from a common ancestor, and this in turn depends on the time their resepective ancestors have spent together in one urn. When we trace back the 'historical process', descibing where ancestors of the two balls lived at previous times, then the difference between their positions follows the symmetrized random walk. If this symmetrized random walk is recurrent, then ancestors of the two balls have for a long time lived together in one urn, and therefore with high proability descend from a common ancestor. This implies that with large probability all balls in one urn are of the same color, and also that after a sufficiently long time any two urns at a finite distance of each other will contain balls of the same color. This explains (1.2.60).

In the article 'Clustering of Linearly Interacting Diffusions and Universality of their Long-Time Distribution' [41], contained in Chapter 3, we show that this picture holds as long as the w-harmonic functions are invariant,[5] and in that case we can even specify the distribution of $X_i(\infty)$ explicitly. In fact, for each $\theta \in K$ there exists a unique probability distribution on $\partial_w K$ with a given harmonic mean. If we call this distribution $\Gamma_\theta(dx)$, then by the fact that the harmonic mean is conserved we see that we must have

$$P[X_i(\infty) \in dx] = \Gamma_\theta(dx). \qquad (1.2.62)$$

The proof that the recurrence of the symmetrized random walk implies $X(t) \Rightarrow X(\infty)$, where $X(\infty)$ satisfies (1.2.60) and (1.2.62), consists of two main ingredients: a calculation of the covariances $\mathrm{Cov}(X_i(t), X_j(t))$ between urns i and j, and a calculation of the expectation of w-harmonic functions.

If we return to the hierarchical group Ω_N in the limit of large N, then modulo the technical difficulties explained in section 1.2.5, we may expect the following

[5] In Chapter 3, we use the terminology 'the boundary distribution is stable against a linear drift'. Under a weak technical assumption, this is equivalent to saying that the w-harmonic functions are invariant; see formulas (3.1.42) and (3.1.46).

behavior. If

$$\sum_{k=1}^{\infty} \frac{1}{c_k} = \infty, \tag{1.2.63}$$

then

$$\left.\begin{array}{l} K_\theta^{w,(k)}(\cdot) \Rightarrow \Gamma_\theta(\cdot) \\ \sigma_k F^{(k)} w \to w^* \end{array}\right\} \quad \text{as } k \to \infty. \tag{1.2.64}$$

Here $\Gamma_\theta(\cdot)$ is the unique distribution on the effective boundary $\partial_w K$ with harmonic means θ, and the fixed shape matrix w^* is given by

$$w^*_{\alpha\beta}(x) := \int_K \Gamma_x(dy)(y_\alpha - x_\alpha)(y_\beta - x_\beta). \tag{1.2.65}$$

Note that the convergence in (1.2.64) is universal in all w with the same invariant w-harmonic functions and in all c_k satisfying (1.2.63). One can check that the random walk with the kernel in (1.2.7) is (for large N) recurrent if and only if (1.2.63) holds.

We end this section by giving a short overview of what we have proved for the p-type q-tuple models in particular.

The first question is: Does the p-type q-tuple model have invariant harmonics? The answer is:

$q \downarrow$ \ $p \rightarrow$	2	3	4	5	6
2	yes	yes	yes	yes	yes
3		yes	no	no	no
4			yes	no	no
5				yes	no
6					yes

I.e., the p-type 2-tuple and the p-type p-tuple models have invariant harmonics, the others have not.

The second question is: If we have invariant harmonics, then is the fixed shape w^* the same as $w^{(p,q)}$ or not? The answer is:

$q \downarrow$ \ $p \rightarrow$	2	3	4	5	6
2	yes	yes	yes	yes	yes
3		yes			
4			no		
5				no	
6					no

I.e., the p-type 2-tuple models and the 3-type 3-tuple model are already in the fixed shape, the rest is not. This means that the p-type 2-tuple models and the 3-type 3-tuple model look the same on different space-time scales, i.e., they are self-similar. But the p-type p-tuple models with $p \geq 4$ look different on large space-time scales than locally.

The third question is: Has uniqueness of solutions to the martingale problem been proved for the operator

$$c \sum_{\alpha} (\theta_\alpha - x_\alpha) \frac{\partial}{\partial x_\alpha} + \sum_{\alpha\beta} w_{\alpha\beta}^{(p,q)}(x) \frac{\partial^2}{\partial x_\alpha \partial x_\beta} \tag{1.2.66}$$

for all $c \in [0, \infty)$ and $\theta \in \hat{K}_p$? The answer is:

$q \downarrow$ \ $p \rightarrow$	2	3	4	5	6
2	yes	yes	yes	yes	yes
3		no	no	no	no
4			no	no	no
5				no	no
6					no

I.e., uniqueness has only been solved completely for the p-type 2-tuple models. However, for $c = 0$ or for c sufficiently large and $\theta \in K^\circ$, uniqueness is known for all p-type q-tuple models.

1.2.8 Doing the iterations at once

After Chapter 3, which shows that systems of linearly interacting diffusions of the form (1.2.55) cluster in a universal way, we return in Chapter 4 to the questions of large space-time scale universality that we first attacked in Chapter 2. Thus, we consider a system of diffusion processes $(X_i^N(t))_{t\geq 0,\, i\in\Omega_N}$ indexed by the N-dimensional hierarchical group, solving a system of stochastic differential equations of the form (1.2.8), where the diffusion matrix is isotropic:

$$\sigma_{\alpha\beta}(x) = \delta_{\alpha\beta}\sqrt{2g(x)}. \tag{1.2.67}$$

As explained in Conjecture 1.2.1, we expect k-block averages of this process to converge, as $N \to \infty$, to a process with a renormalized dynamics:

$$(X_0^{N,k}(N^k t))_{t\geq 0} \Rightarrow (Z_\theta^{F^{(k)}g,c_{k+1}}(t))_{t\geq 0} \qquad \text{as } N \to \infty. \tag{1.2.68}$$

Because of Proposition 1.2.2 and the generalization thereof in section 1.2.5, we expect furthermore that

$$\lim_{k\to\infty} \sup_{x\in K} \left|\sigma_k(F^{(k)}g)(x) - g^*(x)\right| = 0, \tag{1.2.69}$$

where g^* is the function in (1.2.43). It can be shown that (1.2.68) and (1.2.69) together imply that

$$(X_0^{N,k}(\sigma_k N^k t))_{t\geq 0} \Rightarrow (Z_\theta^{g^*,c^*}(t))_{t\geq 0} \qquad \text{as } N \to \infty, k \to \infty, \tag{1.2.70}$$

when $\sigma_k c_{k+1} \to c^*$ for some $c^* \in [0,\infty)$, where the limits have to be taken in the order indicated. However, as we explained in section 1.2.5, a rigorous proof of (1.2.68) is at present out of reach, due to the fact that we have inadequate control over the renormalization transformations F_c.

The idea of Chapter 4 is that we may avoid these problems if we replace the double limit in (1.2.70) by a single one, and try to prove that we can let N and k tend to infinity together in such a way that (1.2.70) holds. Of course, this requires that we are able to understand the universality in (1.2.70) in a way that *does not revert to the renormalization transformations* F_c.

In fact, we can do this by making use of the techniques involving harmonic functions and covariance calculations developed in Chapter 3. The reasons behind the large space-time scale universality in (1.2.70) are, however, more subtle than those behind the universality of the clustering of the process. Clustering, after all, refers only to large time, while for (1.2.70) we must rescale time *and* space.

It appears that large space-time scale universality in the system arises due to the occurrence of 'local equilibrium'. As explained in section 1.2.2, we expect large

k-blocks to reach equilibrium subject to the value of the $(k+1)$-block average. It is from the equilibrium conditions for such a k-block, plus a calculation involving harmonic functions and covariances, that (1.2.70) can be deduced, albeit in a certain weak form. Technically, however, it is not easy to make these arguments work. For fixed N and k, it is not true that the system, or the k-blocks within the system, reach any sort of a non-trivial equilibrium as t tends to infinity, as is clear from our discussion of clustering. Thus, we have to show that, for N and k sufficiently large, k-blocks are sufficiently close to equilibrium for our purposes, on times of an appropriate order of magnitude.

Chapter 4, which is to be part of a forthcoming article, contains work that is still in progress. At the moment, the author has shown that the diffusion rate of large k-block averages tends to g^* when

$$\frac{k}{\log N} \to 0 \qquad \text{as } N \to \infty, k \to \infty. \tag{1.2.71}$$

But the type of convergence shown for the diffusion rate of $X_0^{N,k}$ is not strong enough to conclude that also the process $X_0^{N,k}$ itself converges as in (1.2.70). As the research contained in Chapter 4 is only a few months old, it is at present too early to say if these problems are serious. The impression is, however, that this shortcoming could be inherent in the techniques that are being used. In particular, it seems to be related to the type of convergence to equilibrium that is used. In Chapter 4 it is only shown that k-blocks are approximately in equilibrium when their distribution is averaged over a short time interval. It is conceivable that (1.2.70) can only be derived if this time-averaged equilibrium can be replaced by a (much more difficult to prove) equilibrium at fixed times.

1.3 Open problems

Apart from the obvious open problems concerning the unproved conjectures in (mainly) Chapters 2 and 4, there are a number of questions calling for attention, which we present here.

1.3.1 Non-invariant harmonics

One of the main messages of this dissertation is that the large time universality (clustering) and large space-time universality (renormalization) which have so far been observed in systems of linearly interacting diffusions, find their origin in the fact that the w-harmonic functions associated with these systems are invariant under the semigroup $(T_t^{x,c})_{t\geq 0}$, (that is: (1.2.53) holds). Because of technical difficulties (concerning uniqueness, ergodicity and coupling arguments that are available for one-dimensional K only), most work on linearly interacting diffusions in the past years has focussed on systems with one-dimensional state space K. These systems all have invariant harmonics. But with the study of higher-dimensional state spaces it has become clear that there are many systems, arising naturally as limits of particle models, for which the harmonic functions are not invariant. For example, the p-type q-tuple model with $2 < q < p$ does not have invariant harmonics (see the figure on page 29). A natural question is: do such systems exhibit universal behavior on large time and/or space scales?

As regards the long time behavior, it seems that the answer is no. It seems that invariant harmonics are necessary for the conservation of the harmonic means (formula (1.2.57)), and this in turn seems to be necessary to guarantee that the distribution of $X_i(\infty)$ is universal in the class of all diffusion matrices w with the same w-harmonic functions and in all Abelian groups Λ with a recurrent symmetrized random walk.

Since the results in Chapter 4 indicate a strong link between the large time universality and the large space-time universality, it seems that the same negative conclusion must be drawn about the action of the renormalization transformations F_c on such diffusion matrices w with non-invariant harmonics. In fact, it seems likely that for certain choices of the constants c_k the rescaled iterates $\sigma_k(F_{c_k} \circ \cdots \circ F_{c_1})w$ do not converge to any limit at all. However, this does not exclude the possibility that for certain choices of the c_k we may still find some form of universality.

Consider in particular the case that $c_k = c^k$, where $c \in (0, \infty)$ is some constant. Then we have recurrence of the symmetrized random walk, and hence clustering, iff $c \leq 1$. Because of the scaling relation

$$F_{\lambda c}(g) = \lambda F_c(\tfrac{1}{\lambda}g) \qquad (\lambda \in (0, \infty)), \tag{1.3.1}$$

we can express $\sigma_{k+1}F^{(k+1)}(w)$ in terms of $\sigma_k F^{(k)}(w)$ in the following way:

$$\sigma_{k+1}F^{(k+1)}(w) = \frac{\sigma_{k+1}}{\sigma_k} F_{\sigma_k c^{k+1}}(\sigma_k F^{(k)}(w)). \tag{1.3.2}$$

If $c < 1$, then

$$\sigma_k c^{k+1} = c^{k+1}\sum_{l=1}^{k} c^{-l} \to \frac{c}{1-c} \qquad \text{as } k \to \infty, \tag{1.3.3}$$

but in the *critically recurrent* case $c = 1$ we see that $\sigma_k c^{k+1} \to \infty$ as $k \to \infty$. This means that for large k it is only the large c limit of the renormalization transformation F_c that is important for us. An expansion in c^{-1} gives

$$(F_c w)_{\alpha\beta}(x) = w_{\alpha\beta}(x) + c^{-1}\tfrac{1}{2}\sum_{\gamma\delta} w_{\gamma\delta}(x)\tfrac{\partial^2}{\partial x_\gamma \partial x_\delta} w_{\alpha\beta}(x) + \mathcal{O}(c^{-2}). \tag{1.3.4}$$

Thus, one is tempted to look for 'asymptotic fixed shapes', which would have to solve the equation

$$\sum_{\gamma\delta} w^*_{\gamma\delta}(x) \frac{\partial^2}{\partial x_\gamma \partial x_\delta} w^*_{\alpha\beta}(x) = \lambda w^*_{\alpha\beta}(x) \qquad (x \in K) \tag{1.3.5}$$

for some $\lambda \in (0, \infty)$.

Although all this is rather speculative, it seems that if any form of universality holds for systems with non-invariant harmonics, then we are most likely to find it in the class of critically recurrent systems. Such universality would be interesting, because it would be the first example of universality that does not follow from invariant harmonics.

1.3.2 Renormalization on other lattices

The critically recurrent symmetrized random walk also becomes important if one tries to prove large space-time results for other lattices than the high N limit of the hierarchical group Ω_N. One such result has been derived by Klenke [23]. For interacting diffusions with one-dimensional state space, indexed by the hierarchical lattice Ω_N with *N fixed*, he was able to give a description of the law of large block averages in terms of a Wright-Fisher diffusion process. The important object to look at in this case is the so-called *interaction chain*. This is the chain of all block averages up to a certain size, observed at a given time βt:

$$(X_0^{N,k}(\beta t), \ldots, X_0^{N,0}(\beta t)), \tag{1.3.6}$$

introduced in formula (1.2.15). If one lets N tend to infinity, then in the right time scale β this chain converges to a 'backward' Markov chain, as explained in Conjecture 1.2.1. For finite N, the interaction chain does not have the Markov property, but in the critically recurrent case one can let k tend to infinity with the result that (1.3.6) in the right scaling converges to a diffusion process. These facts are so far known for one-dimensional K only. Their proofs depend on moment calculations involving a dual model, which are not available for the isotropic models treated in this dissertation. It seems worthwhile to investigate if for isotropic models they can be obtained by alternative methods.

1.3.3 Discrete models

In section 1.1 we have seen how certain discrete particle models, closely related to the voter model, have a continuum limit: the diffusion models discussed in this dissertation. If we could rigorously prove the convergence of a given particle model to the associated diffusion model, then our results would make direct contact with the theory of these particle models. This is certainly something worth trying. Apart from this, we can more generally take the results in this dissertation as a motivation to try to prove analogous results for the particle models.

For example, we should expect that the discrete p-type p-tuple model, just like its diffusion counterpart, clusters if and only if the symmetrized random walk is recurrent. Moreover, for infinite Abelian groups Λ (but not for finite Λ!), we expect the large-time frequencies of the remaining $p-1$ colors to be given by the distribution of Brownian motion, starting from θ (with θ as in (1.1.55)), when it first hits the boundary ∂K. However, it seems that a proof of this claim is more difficult for the particle model than for the diffusion model.

We see that although the diffusion models are a lot harder to define than the discrete particle models, because of the use of (rather heavy) diffusion theory, they also make certain things easier.

1.3.4 Outlook and conclusion

We have seen how certain systems of linearly interacting diffusions exhibit universal behavior on large time scales and on large space-time scales. We have come to understand this universality as a phenomenon that is caused by a special property of the systems, which we have called 'invariant harmonics'.

Technical difficulties often forced us to prove weaker theorems than we originally planned, but a nice aspect of the models considered is that they show, so to say, a wide variety in tractability. For the 2-type 2-tuple model, there are coupling techniques available and there is a duality. For general p-type 2-tuple models, the coupling techniques seem to fail, but there is still a duality, while for the p-type p-tuple models there is no duality, but there are calculations involving covariances and harmonic functions that are in some way a substitute. Finally, there are models like the 4-type 3-tuple model, for which we are still hardly able to prove anything.

Chapter 2

Renormalization of Hierarchically Interacting Isotropic Diffusions

Abstract

We study a renormalization transformation arising in an infinite system of interacting diffusions. The components of the system are labeled by the N-dimensional hierarchical lattice ($N \geq 2$) and take values in a compact convex set $\overline{D} \subset \mathbb{R}^d$ ($d \geq 1$). Each component starts at some $\theta \in D$ and is subject to two motions: (1) an isotropic diffusion according to a local diffusion rate $g : \overline{D} \to [0, \infty)$ chosen from an appropriate class; (2) a linear drift towards an average of the surrounding components weighted according to their hierarchical distance. In the local mean-field limit $N \to \infty$, block averages of diffusions within a hierarchical distance k, on an appropriate time scale, are expected to perform a diffusion with local diffusion rate $F^{(k)}g$, where $F^{(k)}g = (F_{c_k} \circ \cdots \circ F_{c_1})g$ is the k-th iterate of *renormalization transformations* F_c ($c > 0$) applied to g. Here the c_k measure the strength of the interaction at hierarchical distance k. We identify F_c and study its orbit $(F^{(k)}g)_{k \geq 0}$. We show that there exists a 'fixed shape' g^* such that $\lim_{k\to\infty} \sigma_k F^{(k)}g = g^*$ for all g, where the σ_k are normalizing constants. In terms of the infinite system, this property means that there is *complete universal behavior* on large space-time scales.

Our results extend earlier work for $d = 1$ and $\overline{D} = [0, 1]$ resp. $[0, \infty)$. The renormalization transformation F_c is defined in terms of the ergodic measure of a d-dimensional diffusion. In $d = 1$ this diffusion allows a Yamada-Watanabe-type coupling, its ergodic measure is reversible and the renormalization transformation

F_c is given by an explicit formula. All this breaks down in $d \geq 2$, which complicates the analysis considerably and forces us to new methods. Part of our results depend on a certain martingale problem being well-posed.

2.1 Introduction

In this paper we study a renormalization transformation that arises in the study of a system of hierarchically interacting diffusions. Our study is part of a larger area where the goal is to understand universal behavior on large space-time scales of stochastic systems with interacting components. In a recent series of papers ([1], [8], [9], [10]), it was shown how renormalization techniques can be used to give a rigorous analysis of a model, described below, consisting of interacting diffusions indexed by the hierarchical lattice and taking values in the state space [0, 1]. In the meantime the analysis has been generalized to the state space $[0, \infty)$ ([2], [11]).

So far, the model has only been treated completely in the case of a one-dimensional state space (although some limited results for the infinite-dimensional state space of probability measures on [0, 1] can be found in [12], [13]). The present paper investigates a class of isotropic models with state space $\overline{D}$, where $D \subset \mathbb{R}^d$ ($d \geq 1$) is open, bounded and convex. To help the reader, we use the remainder of this section to present an overview of the known results for the case $d = 1$, together with a heuristic view on what is behind these results. This overview provides the essential motivation for section 2.2, where we state our new results for the case $d \geq 2$ and formulate some open problems. Proofs appear in sections 2.3–2.5.

2.1.1 Genetic diffusions

Our model finds its origin in population dynamics. Consider a gene that comes in $d + 1$ types ('alleles'). Consider a population consisting of n individuals, each carrying one copy of the gene ('haploid organisms'). At any time the population may be described by a point x in the discrete simplex

$$K_d^n := \{x = (x_1, \ldots, x_d) \in \tfrac{1}{n}\mathbb{Z}^d : x_i \geq 0,\ \sum_{i=1}^{d} x_i \leq 1\}. \tag{2.1.1}$$

We interpret $x_1, \ldots, x_d, 1 - \sum_{i=1}^{d} x_i$ as the proportions of alleles $1, \ldots, d + 1$. Frequencies of alleles are supposed to change due to 'random sampling' and 'migration'.

Random sampling is a random process by which some alleles may occasionally produce more offspring than others. We can model it as a Markov evolution on K_d^n by replacing pairs of individuals after an exponential waiting time with mean 1. A pair is replaced in the following manner: we choose one individual of the pair at random, determine its allele and replace both individuals by individuals with this allele.

Migration is a random process that we can model by introducing a huge reservoir of individuals, with gene frequencies $\theta_1, \ldots, \theta_d$, $1 - \sum_{i=1}^d \theta_i$, and letting each individual in the population be replaced with rate c by an individual of the reservoir.

The generator A of the resulting process (migration and random sampling) is given by

$$\begin{aligned}(A_n f)(x) &= cn \sum_{i,j=1}^{d+1} \theta_i x_j \left[f(x + \frac{e^i}{n} - \frac{e^j}{n}) - f(x) \right] \\ &\quad + n^2 \sum_{i,j=1}^{d+1} x_i x_j \left[f(x + \frac{e^i}{n} - \frac{e^j}{n}) - f(x) \right],\end{aligned} \tag{2.1.2}$$

where $x = (x_1, \ldots, x_d)$ and $e^i = (e^i_1, \ldots, e^i_d)$ with $e^i_j = \delta_{ij}$ for $i = 1, \ldots, d$. In (2.1.2) we have additionally defined $x_{d+1} = 1 - \sum_{i=1}^d x_i$ and $\theta_{d+1} = 1 - \sum_{i=1}^d \theta_i$, and put e^{d+1} to be the zero vector in $\mathbb{R}^d$.

In the limit $n \to \infty$ the gene frequencies take values in the d-dimensional simplex

$$K_d := \{x \in \mathbb{R}^d : x_i \geq 0, \sum_{i=1}^d x_i \leq 1\}. \tag{2.1.3}$$

On functions $f \in \mathcal{C}^2(K_d)$ the generator A_n can be seen to converge, in an appropriate sense, to

$$(Af)(x) = \Big(\sum_{i=1}^d c(\theta_i - x_i) \tfrac{\partial}{\partial x_i} + \sum_{i,j=1}^d x_i(\delta_{ij} - x_j) \tfrac{\partial^2}{\partial x_i \partial x_j} \Big) f(x). \tag{2.1.4}$$

The matrix $x_i(\delta_{ij} - x_j)$ is the Wright-Fisher diffusion matrix. Similar models, with slightly more complicated random sampling mechanisms, yield similar differential operators with different diffusion matrices. Provided the martingale problem for these operators is well-posed, it is often possible to show that the discrete process on K_d^n converges in law to the diffusion with generator A (see [16] for details).

Let us consider the case $d = 1$ and let us introduce the following objects.

1. ('state space') $K_1 = [0, 1]$.

2. ('diffusion function') $\mathcal{H}_{Lip}$ is the class of functions $g : [0, 1] \to [0, \infty)$ satisfying

$$\begin{aligned} &\text{(i)} && g = 0 \text{ on } \{0, 1\} \\ &\text{(ii)} && g > 0 \text{ on } (0, 1) \\ &\text{(iii)} && g \text{ is Lipschitz continuous on } [0, 1]. \end{aligned} \tag{2.1.5}$$

3. ('attraction point') $\theta \in [0, 1]$.

4. ('attraction constant') $c \in (0, \infty)$.

With these ingredients we consider the following Stochastic Differential Equation (SDE) on $[0, 1]$:

$$dX_t = c(\theta - X_t)dt + \sqrt{2g(X_t)}dB_t \qquad (t \geq 0), \tag{2.1.6}$$

where $(B_t)_{t\geq 0}$ is standard Brownian motion. We call (2.1.6) 'the basic diffusion equation'. We define a linear operator A with domain $\mathcal{C}^2[0, 1]$ (the two times continuously differentiable real functions on $[0, 1]$) by putting

$$(Af)(x) := [c(\theta - x)\tfrac{\partial}{\partial x} + g(x)\tfrac{\partial^2}{\partial x^2}]f(x). \tag{2.1.7}$$

The following is known[1]:

Theorem 2.1.1 *For each $g \in \mathcal{H}_{Lip}$, $\theta \in [0, 1]$ and $c \in [0, \infty)$, and for each initial distribution on $[0, 1]$, the SDE (2.1.6) has a unique strong solution $(X_t)_{t\geq 0}$. The martingale problem for A in (2.1.7) is well-posed and the law of $(X_t)_{t\geq 0}$ solves the martingale problem for A. The operator A has a unique extension to a generator of a Feller semigroup and $(X_t)_{t\geq 0}$ is the associated Feller process.*

The choice $g(x) = x(1 - x)$ corresponds to the Wright-Fisher case. The diffusion equation (2.1.6) and its generalizations to higher dimensions will play a key role in the present paper. In the following sections we show how it has been used as a starting point for the construction and analysis of an infinite system of interacting diffusions.

[1]To get these results, extend the diffusion function g by putting $g(x) = g(1)$ $(x \geq 1)$, $g(x) = g(0)$ $(x \leq 0)$, and extend the drift by the same recipe. It is easy to show that any solution $(X_t)_{t\geq 0}$ of the SDE on $\mathbb{R}$ satisfies $P[X_t \in [0, 1]\ \forall t \geq 0] = 1$. Now, by Skorohod's Theorem ([22], Theorem 5.4.22), there exists a weak solution of (2.1.6). The Yamada-Watanabe argument ([22], Proposition 5.2.13) gives strong uniqueness, and therefore strong existence as well as weak uniqueness ([22], Propositions 5.3.23 and 5.3.20). It follows that the martingale-problem is well-posed ([22], Corollary 5.4.8 and 5.4.9). The process $(X_t)_{t\geq 0}$ has the Feller property ([40], Corollary 11.1.5) and its generator G clearly extends A. In fact, it is the only generator of a Feller semigroup to do so. For let $\tilde{G}$ be another generator extending A, then there exists an associated Feller process $(\tilde{X}_t)_{t\geq 0}$ ([16], Theorem 4.2.7) that solves the martingale problem for A ([16], Theorem 4.1.7). It follows that $(\tilde{X}_t)_{t\geq 0}$ and $(X_t)_{t\geq 0}$ have the same distribution for all initial conditions, and hence $G = \tilde{G}$.

In the special case that $g \in \mathcal{C}^2[0, 1]$, it is known that G is the closure of A ([16], Theorem 8.2.1), but for general $g \in \mathcal{H}_{Lip}$ this seems to be an open problem. (In this respect, the loose remark in [1], page 7, that the closure of the operator G mentioned there generates a Feller semigroup seems unfounded.)

2.1.2 The hierarchical model

In the model described in the previous section, all individuals have equal chances of interaction with all other individuals. A more realistic model takes into account the effects of isolation by distance. To this aim, we introduce the following additional objects:

5. ('index space') For $N \geq 2$, let Ω_N be the N-dimensional hierarchical lattice

$$\Omega_N := \Big\{(\xi_i)_{i\geq 1} : \xi_i \in \{0, 1, \dots, N-1\}, \xi_i \neq 0 \text{ finitely often}\Big\}. \qquad (2.1.8)$$

   With componentwise addition (mod N), Ω_N is a countable group.

6. ('distance') Let $d : \Omega_N \times \Omega_N \to \mathbb{N}_0$ be the hierarchical distance

$$d(\xi, \eta) := \min\{j \geq 0 : \xi_i = \eta_i \text{ for all } i > j\}. \qquad (2.1.9)$$

7. ('interaction constants') Let $(c_k)_{k\geq 1}$ be strictly positive constants, satisfying

$$\sum_{k=1}^{\infty} c_k^{-1} = \infty \qquad (2.1.10)$$

$$\sum_{k=1}^{\infty} c_k N^{-k} < \infty. \qquad (2.1.11)$$

8. ('noise') Let $(\{B_\xi(t)\}_{\xi\in\Omega_N})_{t\geq 0}$ be an i.i.d. collection of standard Brownian motions.

With the above ingredients, we consider the process

$$X^N = \Big(X^N(t)\Big)_{t\geq 0} = \Big(\Big\{X_\xi^N(t)\Big\}_{\xi\in\Omega_N}\Big)_{t\geq 0} \qquad (2.1.12)$$

with state space $[0, 1]^{\Omega_N}$ given by the following set of coupled SDE's:

$$\begin{aligned} dX_\xi^N(t) &= \textstyle\sum_{k=1}^{\infty} c_k N^{1-k}\Big[X_\xi^{N,k}(t) - X_\xi^N(t)\Big]dt + \sqrt{2g\Big(X_\xi^N(t)\Big)}dB_\xi(t) \\ X_\xi^N(0) &= \theta \qquad (t \geq 0, \xi \in \Omega_N), \end{aligned} \qquad (2.1.13)$$

where $X_\xi^{N,k}(t)$ is the block average

$$X_\xi^{N,k}(t) := \frac{1}{N^k} \sum_{\eta:\, d(\eta,\xi)\leq k} X_\eta^N(t) \qquad (k = 0, 1, 2\dots). \qquad (2.1.14)$$

The system in (2.1.13) can be interpreted as a model for the time evolution of gene distributions in an infinite population (see [8], [33] for the origin of the model and [16], Chapter 10, for more background). The population is organized in sites, groups, clans, villages etc., where N sites form a group, N groups form a clan, N clans form a village, and so on. The index space Ω_N labels sites by numbering sites within a group by a number $\xi_1 = 0, \ldots, N-1$, numbering groups within a clan by a number $\xi_2 = 0, \ldots, N-1$, and so on. (For example, if the distance between two sites ξ and η is $d(\xi,\eta) = 2$, then ξ and η share the same village and clan but belong to different groups.) The proportion of allele 1 at a given site ξ at some time t is described by $X^N_\xi(t)$. Initially, all proportions are supposed to be θ, and they evolve due to migration and random sampling, as in (2.1.6). However, the migration is now supposed to cause interaction between sites mutually (see (2.1.16–2.1.17) below), instead of between sites and some infinite reservoir. The meaning of the numbers $X^{N,k}_\xi(t)$ is the following: $X^{N,0}_\xi(t) = X^N_\xi(t)$ is the proportion of individuals of allele 1 at site ξ; $X^{N,1}_\xi(t)$ is the proportion in the group that ξ belongs to; $X^{N,2}_\xi(t)$ is the proportion in the clan that ξ belongs to, and so on. We call the set $\{\eta : d(\eta,\xi) \leq k\}$ the 'k-block' around ξ and the numbers $X^{N,k}_\xi(t)$ the 'k-block averages' around ξ.

The factor $c_k N^{1-k}$ in (2.1.13) describes the strength of the interaction (due to migration) between a site ξ and the k-block around ξ. The strength of the attraction decays by a factor $1/N$ each time we go up one step in the hierarchy. As we shall see later, precisely this decay will give rise to non-trivial behavior in the limit as $N \to \infty$.

The next theorem follows from [39], Theorem 3.2:

Theorem 2.1.2 *Let $N \geq 2$, $g \in \mathcal{H}_{Lip}$, $c_k \in [0,\infty)$ $(k \geq 1)$, $\sum_k c_k N^{-k} < \infty$ and $\theta \in [0,1]$. Then the system of SDE's in (2.1.13) has a unique strong solution satisfying*

$$P[X^N_\xi(t) \in [0,1] \quad \forall\, \xi \in \Omega_N,\ t \geq 0] = 1. \tag{2.1.15}$$

We need to check that [39], Assumption [B-2]$'$ is satisfied. The drift term in (2.1.13) can be rewritten as

$$\sum_{k=1}^{\infty} c_k N^{1-k}\Big[X^{N,k}_\xi(t) - X^N_\xi(t)\Big]dt = \sum_{\eta\in\Omega_N} a_N(\xi,\eta)\Big[X^N_\eta(t) - X^N_\xi(t)\Big]dt, \tag{2.1.16}$$

where

$$a_N(\xi,\eta) = \sum_{k=d(\xi,\eta)}^{\infty} c_k N^{1-2k}. \tag{2.1.17}$$

Hence the drift term is in fact a pair interaction between the different components. A little calculation shows that $\sum_{\eta\in\Omega_N} a_N(\xi,\eta) = \sum_{k=1}^{\infty} c_k N^{1-k}\ \forall \xi \in \Omega_N$. Condition (2.1.11) is therefore exactly what is required in [39], Assumption [B-2]′.

2.1.3 The local mean-field limit $N \to \infty$

We shall study the system in (2.1.13) in the limit as $N \to \infty$. The 1-block average $X_\xi^{N,1}(t)$ is the average of a large number of diffusions that behave independently apart from their linear drift towards block averages.

Let Δt be small and let $\Delta X_\xi^N(t) := X_\xi^N(t+\Delta t) - X_\xi^N(t)$. Let $\mathcal{F}_t$ be the σ-field of events up to time t. Then (2.1.13) can, in a heuristic way, be rewritten as

$$\begin{aligned} E[\Delta X_\xi^N(t) \mid \mathcal{F}_t] &\cong \textstyle\sum_{k=1}^{\infty} c_k N^{1-k}\Big[X_\xi^{N,k}(t) - X_\xi^N(t)\Big]\Delta t \\ E[\Delta X_\xi^N(t)\Delta X_\eta^N(t) \mid \mathcal{F}_t] &\cong \delta_{\xi,\eta} 2g(X_\xi^N(t))\Delta t. \end{aligned} \tag{2.1.18}$$

It follows that for the 1-block averages we have

$$\begin{aligned} E[\Delta X_\xi^{N,1}(t) \mid \mathcal{F}_t] &\cong \textstyle\sum_{k=2}^{\infty} c_k N^{1-k}\Big[X_\xi^{N,k}(t) - X_\xi^{N,1}(t)\Big]\Delta t \\ E[\Delta X_\xi^{N,1}(t)\Delta X_\eta^{N,1}(t) \mid \mathcal{F}_t] &\cong 1_{\{d(\xi,\eta)\le 1\}} 2N^{-2} \textstyle\sum_{\zeta:\, d(\zeta,\xi)\le 1} g(X_\zeta^N(t))\Delta t. \end{aligned} \tag{2.1.19}$$

Note that in the first line the term with $k=1$ drops out. Note further that in the second line the sum $\sum_{\zeta:\, d(\zeta,\xi)\le 1}$ is over N terms. Hence both expectations are of order N^{-1}. We are therefore led to believe that the 1-block average $X_\xi^{N,1}(t)$ moves slowly w.r.t. $X_\xi^N(t)$, namely, its time scale is Nt rather than t. For large N this means that $X_\xi^{N,1}(t)$ stays essentially fixed at its initial value θ. Inserting this into (2.1.13) and neglecting terms of order $1/N$, we see that the single componentscomponent, single $X_\xi^N(t)$ satisfy a limiting SDE of the type (2.1.6) with $c = c_1$. The limit $N \to \infty$ thus corresponds to a 'local mean-field' limit. On the local space scale of 1-blocks, the interaction reduces to a linear drift towards an essentially fixed block average, so that the single components are asymptotically independent (in physics language: the system shows 'propagation of chaos'). This behavior, however, occurs only locally. We shall see later that on larger space scales the interaction still gives rise to nontrivial correlations between components.

A detailed study of the basic diffusion equation (2.1.6) is the key to understanding the system in (2.1.13). In particular, the invariant measure of (2.1.6) plays a key role. The following theorem is generally known (it can be proved using the coupling mentioned in (2.2.23)).

Theorem 2.1.3 *For every $g \in \mathcal{H}_{Lip}$, $\theta \in [0,1]$ and $c \in (0,\infty)$, the SDE in (2.1.6) has a unique equilibrium $\nu_\theta^{g,c}$ and is ergodic, i.e., for any $x \in [0,1]$ the law of X_t given $X_0 = x$ converges weakly to $\nu_\theta^{g,c}$ as $t \to \infty$. The measure $\nu_\theta^{g,c}$ is given by*

$$\begin{aligned} \nu_\theta^{g,c}(dx) &= \frac{1}{Z_\theta^{g,c}} \frac{1}{g(x)} \exp\left(\int_\theta^x \frac{y-\theta}{g(y)} dy\right) dx \qquad && (\theta \in (0,1)) \\ \nu_\theta^{g,c}(dx) &= \delta_\theta(dx) && (\theta \in \{0,1\}), \end{aligned} \tag{2.1.20}$$

where $Z_\theta^{g,c}$ is a normalization constant depending on g, c and θ.

For $\theta \in (0,1)$, the density of $\nu_\theta^{g,c}$ solves the equation $(c(x-\theta) + \frac{\partial}{\partial x} g(x))\nu_\theta^{g,c}(x) = 0$ (compare (2.2.25) and (2.3.17) (ii)).

2.1.4 The renormalization transformation

The reasoning above indicates that, for large N, the single components $X_\xi^N(t)$ perform a diffusion as in (2.1.6), with as a stochastic attraction point the 1-block average $X_\xi^{N,1}(t)$. Since the single components reach equilibrium on time scale t (i.e., fast compared to time scale Nt of the block), we expect that at times of order Nt their conditional distribution given the 1-block average is given by

$$P[X_\xi^N(Nt) \in dy \mid X_\xi^{N,1}(Nt) = x] \cong \nu_x^{g,c_1}(dy). \tag{2.1.21}$$

Now again consider the heuristic formula (2.1.19). Formula (2.1.21) suggests that

$$N^{-1} \sum_{\zeta:\, d(\zeta,\xi)\le 1} g(X_\zeta^N(Nt)) \cong \int_{[0,1]} g(y) \nu^{g,c_1}_{X_\xi^{N,1}(Nt)}(dy). \tag{2.1.22}$$

This motivates the following definition of our renormalization transformation: for every $g \in \mathcal{H}_{Lip}$, $c \in (0,\infty)$

$$(F_c g)(x) := \int_{[0,1]} g(z) \nu_x^{g,c}(dz) \qquad (x \in [0,1]). \tag{2.1.23}$$

From [9], Lemma 2.2 it follows that:

Theorem 2.1.4 *For all $c \in (0,\infty)$: $F_c \mathcal{H}_{Lip} \subset \mathcal{H}_{Lip}$.*

Theorem 2.1.4 makes it possible to speak about the iterates of F_c, which we shall need below.

2.1.5 Multiple space-time scale analysis

Combining (2.1.19) with (2.1.22) and (2.1.23), and neglecting higher order terms in N, we find the following conditional expectations for $X_\xi^{N,1}(t)$:

$$\begin{aligned} E[\Delta X_\xi^{N,1}(t) \mid \mathcal{F}_t] &\cong N^{-1}c_2\Big[X_\xi^{N,2}(t) - X_\xi^{N,1}(t)\Big]\Delta t \\ E[\Delta X_\xi^{N,1} \Delta X_\eta^{N,1} \mid \mathcal{F}_t] &\cong N^{-1}1_{\{d(\xi,\eta)\le 1\}}2(F_{c_1}g)(X_\xi^{N,1}(t))\Delta t. \end{aligned} \tag{2.1.24}$$

Note that $1_{\{d(\xi,\eta)\le 1\}} = 1$ if and only if the 1-block around ξ is the 1-block around η. The conditional expectations above seem to indicate that 1-block averages, when viewed on time scale Nt, behave as diffusions like the single components, *but with the local diffusion rate g replaced by* $F_{c_1}g$. This is precisely what is proved in [10]. In fact, the reasoning can be extended to arbitrary k-blocks. The local diffusion rate is then $(F_{c_k} \circ \cdots \circ F_{c_1})g$. The time scale for the k-blocks turns out to be $N^k t$. Indeed, we must rescale space and time together: each time we go up one step in the hierarchy we have larger blocks moving on a slower time scale.

To be precise, the heuristic formula (2.1.21) is justified for general k by the following theorem ([10], Theorem 1). Here, for each N, we take $0 = (0, 0, \ldots) \in \Omega_N$ as a typical reference point, and we denote weak convergence by $\Longrightarrow$.

Theorem 2.1.5 *Fix $g \in \mathcal{H}_{Lip}$, $\theta \in [0, 1]$, $t > 0$ and $k \ge 0$. Then as $N \to \infty$*

$$(X_0^{N,k}(N^k t), \ldots, X_0^{N,0}(N^k t)) \Longrightarrow (Z_k, \ldots, Z_0), \tag{2.1.25}$$

where $(Z_k, \ldots, Z_0)$ is a 'backward' time-inhomogeneous Markov chain with transition kernels

$$P[Z_{l-1} \in dy | Z_l = x] = \nu_x^{F^{(l-1)}g, c_l}(dy) \qquad (l = k, \ldots, 1), \tag{2.1.26}$$

and $F^{(k)}g := (F_{c_k} \circ \cdots \circ F_{c_1})g$ is the k-th iterate of the renormalization transformations F_c applied to g ($F^{(0)}g = g$).

The joint distribution of the $(Z_k, \ldots, Z_0)$ above is determined by the 'backward' transition probabilities in (2.1.26) and the distribution of Z_k. The latter depends on t and can be read off from the next theorem ([10], Theorem 1). Here the $\Longrightarrow$ denotes weak convergence in path space $\mathcal{C}[0, \infty)$.

Theorem 2.1.6 *Fix $g \in \mathcal{H}_{Lip}$, $\theta \in [0, 1]$ and $k \ge 0$. Then as $N \to \infty$*

$$\Big(X_0^{N,k}(N^k t)\Big)_{t\ge 0} \Longrightarrow \Big(Z_\theta^{F^{(k)}g, c_{k+1}}(t)\Big)_{t\ge 0}, \tag{2.1.27}$$

where $(Z_\theta^{g,c}(t))_{t\geq 0}$ *is the unique strong solution of the single component SDE on* $[0,1]$ *given by*

$$\begin{aligned} dZ(t) &= c(\theta - Z(t))dt + \sqrt{2g(Z(t))}dB(t) \\ Z(0) &= \theta. \end{aligned} \tag{2.1.28}$$

For $k=0$ this result justifies our heuristic belief that the single components follow the basic diffusion equation (2.1.6), and for $k=1$ it justifies our formula (2.1.24). For general $k\geq 1$ it describes the behavior of the k-block averages.

As a side remark, we note that the initial condition $X_\xi^N(0)=\theta$ in (2.1.13) can be generalized considerably. In [8], section 2, and [10], Remark below equation (1.5), $\{X_\xi^N(0)\}_{\xi\in\Omega_N}$ is taken to be distributed according to a homogeneous ergodic measure μ with $E^\mu(X_\xi^N(0))=\theta$ for all $\xi\in\Omega_N$. For instance, one can take the $X_\xi^N(0)$ to be i.i.d. with mean θ. In this case, Theorem 2.1.6 changes, in the sense that the distribution of $Z_\theta^{g,c_1}(0)$ is given by μ rather than δ_θ. The distribution of $Z_\theta^{F^{(k)}g,c_{k+1}}(0)$ for $k\geq 1$ is, however, still δ_θ. In view of this, the model where each component starts in θ is the most natural one.

2.1.6 Large space-time behavior and universality

Theorems 2.1.5 and 2.1.6 describe the behavior of our system in the limit as $N\to\infty$. We next study the system by taking one more limit, namely, we consider k-blocks with $k\to\infty$. This gives rise to two more theorems: Theorem 2.1.7 describes the behavior of the Markov chain in Theorem 2.1.5 for large k, while Theorem 2.1.9 describes the behavior of the renormalized diffusion function in Theorem 2.1.6 for large k. The translation of these theorems in terms of the infinite system is described in Theorems 2.1.8 and 2.1.10.

As a joint function of θ and dx, the equilibrium $\nu_\theta^{g,c}(dx)$ in (2.1.20) is a continuous probability kernel on $[0,1]$. Let $\mathcal{P}[0,1]$ denote the probability measures on $[0,1]$, equipped with the topology of weak convergence, and let $\mathcal{K}[0,1]$ denote the space of all continuous kernels $K:[0,1]\to\mathcal{P}[0,1]$, equipped with the topology of uniform convergence (see also section 2.2.3). A kernel K evaluated in a point x is denoted by K_x. Uniform convergence of probablitity kernels implies pointwise convergence, so $K^n\to K$ in the topology on $\mathcal{K}[0,1]$ implies $K_x^n\Rightarrow K_x$ for all $x\in[0,1]$.

We denote the composition of two probability kernels $K_x(dy)$ and $L_x(dy)$ by

$$(KL)_x(dz) := \int_{[0,1]} K_x(dy)L_y(dz). \tag{2.1.29}$$

By Theorem 2.1.5, in the limit as $N \to \infty$, the conditional probability of $X_\xi^N(N^k t) \in dy$ given $X_\xi^{N,k}(N^k t) = x$ is given by the kernel

$$K_x^{g,(k)}(dy) := (\nu^{F^{(k-1)}g,c_k} \cdots \nu^{g,c_1})_x(dy), \tag{2.1.30}$$

with the composition as in (2.1.29). The following can be found in [1], equation (1.7):

Theorem 2.1.7 *Fix $g \in \mathcal{H}_{Lip}$. As $k \to \infty$, then in the sense of uniform convergence of probability kernels:*

$$K^{g,(k)} \to K^{(\infty)}, \tag{2.1.31}$$

where the limiting kernel $K^{(\infty)}$ is universal in g and given by

$$K_\theta^{(\infty)} = (1-\theta)\delta_0 + \theta\delta_1 \qquad (\theta \in [0,1]). \tag{2.1.32}$$

Note that, for any $k \geq l$, the conditional probability of $X_\xi^{N,l}(N^k t) \in dy$ given $X_\xi^{N,k}(N^k t) = x$ is described by the kernel $\nu^{F^{(k-1)}g,c_k} \cdots \nu^{F^{(l)}g,c_{l+1}}$, which is just the kernel in (2.1.30) with g replaced by $F^{(l)}g$ and $(c_k)_{k\geq 1}$ replaced by $(c_k)_{k\geq l+1}$. Using Theorem 2.1.5 and the fact that, with $Z(t)$ as in (2.1.28), we have $E[Z(t)] = \theta$ $\forall t \geq 0$, Theorem 2.1.7 translates into the following statement about the infinite system:

Theorem 2.1.8 *Fix $g \in \mathcal{H}_{Lip}$, $\theta \in [0,1]$, $l \geq 0$ and $t > 0$. Then, in the sense of convergence in law:*

$$\lim_{k\to\infty}\lim_{N\to\infty} X_0^{N,l}(N^k t) = Y, \tag{2.1.33}$$

where the law of Y is given by $\mathcal{L}(Y) = (1-\theta)\delta_0 + \theta\delta_1$.

Thus, the system locally ends up in one of the traps 0 or 1. This behavior is called *clustering* and should be interpreted as saying that, for large N and k, the block averages spend most of their time close to the boundaries of the state space [0, 1]. Condition (2.1.10) in fact characterizes the clustering regime for the system in the $N \to \infty$ limit. For finite N, clustering of the system can be related to the recurrence of the random walk with kernel $a_N(\xi,\eta)$ given in (2.1.17) (see [6]). For a discussion of clustering in the case $g(x) = rx(1-x)$, both for $N \to \infty$ and for finite N, see [10], Theorems 3 and 6.

We next turn to the behavior of $F^{(k)}g$ as $k \to \infty$. Note that since $\nu_\theta^{g,c}$ itself depends on g, the transformation F_c is a non-linear integral transform. As such it is a rather difficult object to study in detail. Nevertheless, [1] gives a complete description of the asymptotic behavior of its iterates. The results show that there is a unique 'fixed shape' $g^* \in \mathcal{H}_{Lip}$ that attracts all orbits after appropriate scaling, as follows:

Theorem 2.1.9
(a) *Let* $g^*(x) = x(1-x)$. *The 1-parameter family of functions* $g = rg^*$ $(r > 0)$ *are fixed shapes under* F_c*:*

$$F_c(rg^*) = \left(\frac{c}{c+r}\right) rg^*. \tag{2.1.34}$$

(b) *For all* $g \in \mathcal{H}_{Lip}$

$$\lim_{k\to\infty} \sigma_k F^{(k)} g = g^* \quad \textit{uniformly on } [0,1], \tag{2.1.35}$$

where $\sigma_k := \sum_{l=1}^{k} c_l^{-1}$.
(c) *Let*

$$\mathcal{H}_1 := \{g \in \mathcal{H}_{Lip} : \liminf_{x\to 0} x^{-2} g(x) > 0 \textit{ and } \liminf_{x\to 1} (1-x)^{-2} g(x) > 0\}. \tag{2.1.36}$$

Then for all $g \in \mathcal{H}_1$

$$\lim_{k\to\infty} \|\sigma_k F^{(k)} g - g^*\|_{\mathcal{H}_{Lip}} = 0, \tag{2.1.37}$$

where

$$\|g\|_{\mathcal{H}_{Lip}} := \sup_{x\in(0,1)} \left|\frac{g(x)}{g^*(x)}\right|. \tag{2.1.38}$$

To be able to state the implications of Theorem 2.1.9 for the infinite system, we must rescale the time once more, now to compensate not for the large N but for the large k. Indeed, by an easy scaling property of the $Z_\theta^{g,c}$ defined in (2.1.28), we can rewrite Theorem 2.1.6 as

$$\left(X_0^{N,k}(\sigma_k N^k t)\right)_{t\geq 0} \Longrightarrow \left(Z_\theta^{\sigma_k F^{(k)} g, \sigma_k c_k}(t)\right)_{t\geq 0} \qquad \text{as } N\to\infty. \tag{2.1.39}$$

In view of (2.1.35), the most interesting behavior now occurs when $\sigma_k c_k$ tends to some limit as $k\to\infty$. From Theorem 2.1.9 (b) we get, by a simple application of [40], Theorem 11.1.4, the following:

Theorem 2.1.10 *If* $\lim_{k\to\infty} \sigma_k c_k = c^* \in [0,\infty)$, *then in the sense of weak convergence of the law in path space* $\mathcal{C}[0,\infty)$*:*

$$\lim_{k\to\infty}\lim_{N\to\infty} \left(X_0^{N,k}(\sigma_k N^k t)\right)_{t\geq 0} = \left(Z_\theta^{g^*,c^*}(t)\right)_{t\geq 0}. \tag{2.1.40}$$

For example, if $c_k = ab^k$ with $a \in (0,\infty)$ and $b \in (0,1)$, then $\lim_{k\to\infty} \sigma_k c_k = a^2 \frac{b}{1-b}$.

The results in Theorems 2.1.9 and 2.1.10 show that our system displays *complete universality* on large space-time scales. For large k (and in the limit as

$N \to \infty$) the k-blocks approximately perform the diffusion in (2.1.28) with diffusion function g^* and with attraction constant c^*, *and this behavior is completely universal in the diffusion function g of the single components*.

Theorem 2.1.9 (c) is important for the study of how clustering occurs. In fact, under (2.1.37) *the clustering turns out to be universal in g* (see [10], Corollary at Theorem 5). It turns out that the class $\mathcal{H}_1$ in (2.1.36) is sharp: if $\limsup_{x\to 0} x^{-2}g(x) = 0$ or $\limsup_{x\to 0}(1-x)^{-2}g(x) = 0$, then $\sigma_k F^{(k)}g$ does not converge in the norm $\|\cdot\|_{\mathcal{H}_{Lip}}$ (see [1]).

2.2 Results for $d \geq 1$

In this section we present our best results towards extending the model in section 2.1 to higher dimension. In sections 2.2.1 and 2.2.2 we formulate a general program, and specify the particular model that is the subject of the present paper. In section 2.2.3 we present our theorems on the renormalization transformations F_c ($c \in (0,\infty)$) arising in that model. The theorems are stated in terms of certain classes of functions $\mathcal{H}'$ and $\mathcal{H}''$. These are essentially the largest domains on which we can define our renormalization transformations F_c, resp. the iterates $F^{(k)}$. For the results to make sense, it remains to be shown that these classes are not empty. This task is, with limited success, taken up in section 2.2.4. In section 2.2.5 we indicate some of the difficulties that make life hard in $d \geq 2$. Finally, in section 2.2.6, some of the more urgent open problems are discussed.

Proofs are given in sections 2.3–2.5.

2.2.1 Generalizations to different state spaces

The renormalization techniques described in the last section are not restricted to models with state space $[0,1]$. The construction of more general models could be described in the form of the following program:

1. Choose an open convex domain $D \subset \mathbb{R}^d$ and a class $\mathcal{H}$ of diffusion matrices on $\overline{D}$ (i.e., the equivalents of $[0,1]$ and $\mathcal{H}_{Lip}$ in section 2.1). Prove (as in Theorem 2.1.1) that for all $g \in \mathcal{H}$, $\theta \in \overline{D}$, $c \in (0,\infty)$ the martingale problem is well-posed for the differential operator

$$(Af)(x) := \Big(\sum_{i=1}^{d} c(\theta_i - x_i)\tfrac{\partial}{\partial x_i} + \sum_{i,j=1}^{d} g_{ij}(x)\tfrac{\partial^2}{\partial x_i \partial x_j}\Big)f(x). \qquad (2.2.1)$$

2. Prove (as in Theorem 2.1.3) ergodicity for the diffusion given by A, and define a renormalization transformation F_c by

$$(F_c g)_{ij}(\theta) = \int_{\overline{D}} \nu_\theta^{g,c}(dx) g_{ij}(x), \tag{2.2.2}$$

where $\nu_\theta^{g,c}$ is the equilibrium associated with (2.2.1). Show that $\mathcal{H}$ is closed under F_c (as in Theorem 2.1.4), and show that the iterates $F^{(k)}g$ and the associated kernel $K^{g,(k)}$ describe the multiple space-time scale behavior of the associated infinite system (i.e., prove analogues of Theorems 2.1.2, 2.1.5 and 2.1.6).

3. Investigate the limiting behavior of $K_\theta^{g,(k)}$ and $F^{(k)}g$ as $k \to \infty$ (i.e., try to prove equivalents of Theorems 2.1.7–2.1.10).

So far, such a program has only been carried out completely for one-dimensional state spaces, as explained in section 2.1. For the program to get off the ground, one must at least be able to speak about the iterates $F^{(k)}g$. In practice, this leads to conflicting demands on the class $\mathcal{H}$. When $\mathcal{H}$ is chosen large, it turns out to be difficult to show uniqueness for the martingale problem for A in (2.2.1), and therefore the program already stops at step 1. On the other hand, when $\mathcal{H}$ is chosen too restrictive, it turns out to be hard to show (in step 2) that $F_c g \in \mathcal{H}$, i.e., we can define $F_c g$ but not its iterates $F^{(k)}g$. At present, these difficulties present a serious obstacle in trying to carry out the program above completely for state spaces in dimensions $d \geq 2$.

In the present paper, we focus on the construction of $F^{(k)}g$ and $K^{g,(k)}$ and the study of their limiting behavior for a certain class $\mathcal{H}$ of 'isotropic' diffusion matrices in dimensions $d \geq 2$. We leave the study of the associated infinite system to be treated in future work. The difficulties mentioned above are dealt with in the following way. We introduce subclasses $\mathcal{H}'$ and $\mathcal{H}''$ that are essentially the largest subsets of $\mathcal{H}$ on which $F_c g$ resp. $F^{(k)}g$ can be defined. (It may be that $\mathcal{H} = \mathcal{H}' = \mathcal{H}''$, but this can at present not be proved.) In section 2.2.3, we show that on these classes it is possible to carry out step 3 of the above program completely. In particular, we show that there exists a unique fixed shape g^* under F_c that attracts all g under appropriate scaling, and also, that there exists a universal limiting kernel to which all $K^{g,(k)}$ converge. In section 2.2.4, we investigate under what conditions $F_c g$ and $F^{(k)}g$ can be defined, i.e., we find conditions for $g \in \mathcal{H}'$ and $g \in \mathcal{H}''$. The results in this section are not as conclusive as those in section 2.2.3, but we can show that many functions are in $\mathcal{H}'$, and at least in one example we can show that $\mathcal{H}''$ is not empty.

2.2.2 Isotropic models

We consider as state space the closure $\overline{D}$ of an arbitrary open, bounded and convex set $D \subset \mathbb{R}^d$. On $\overline{D}$, we consider a class $\mathcal{H}$ of isotropic diffusion matrices. We say that a diffusion matrix $g_{ij}(x)$ is isotropic if it has the form $g_{ij}(x) = \delta_{ij} g(x)$, where $g : \overline{D} \to [0, \infty)$ is some non-negative function, and $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ otherwise. From the form of the renormalization transformation we see that for an isotropic diffusion matrix

$$(F_c g)_{ij}(\theta) = \int_{\overline{D}} \nu_\theta^{g,c}(dx) \delta_{ij} g(x) = \delta_{ij} \int_{\overline{D}} \nu_\theta^{g,c}(dx) g(x), \tag{2.2.3}$$

so if g is isotropic, then $F_c g$ is isotropic. On the class of isotropic diffusions, F_c is essentially just a transformation of functions $g : \overline{D} \to [0, \infty)$.

In the special case that $\overline{D} = K_d$, the d-dimensional simplex, we indicate briefly how such isotropic models can arise as continuous limits of discrete models. Consider the random sampling procedure described in section 2.1.1. Suppose that instead of replacing *pairs* we replace $(d + 1)$-*tuples*, in the following manner. After an exponential time, a $(d + 1)$-tuple of individuals is selected. If all $d + 1$ individuals belong to different types, then they are all replaced by one randomly chosen type. Otherwise nothing happens. A little calculation shows that this procedure gives rise to the following diffusion matrix:

$$g_{ij}(x) = [\delta_{ij}(d + 1) - 1] \Big(1 - \sum_k x_k\Big) \prod_k x_k \qquad (i, j, k = 1, \ldots, d). \tag{2.2.4}$$

By a simple transformation of the state space, the matrix $\delta_{ij}(d+1)-1$ can be diagonalized to δ_{ij}. In this way one arrives at an isotropic model with $g_{ij}(x) = \delta_{ij} g(x)$, where g is given by (the transformed function of) $x \mapsto (1 - \sum_k x_k) \prod_k x_k$. More general functions g can be obtained by making the rate of the random sampling process dependent on the state of the system.[2,3]

2.2.3 Renormalization in $d \geq 1$: Theorems 2.2.1–2.2.4

We introduce the following objects:

[2] In section 2.1.1, the Wright-Fisher diffusion was introduced on K_d. In dimensions $d \geq 2$, this diffusion is non-isotropic. It is not hard to see that it is a fixed shape under F_c. Therefore it is expected that in $d \geq 2$, and on a larger class than only the isotropic diffusions treated in the present paper, the transformation F_c has many different fixed shapes, each with their own domain of attraction.

[3] In $d = 1$, the fixed shape g^* on the simplex appears to be the most natural object when seen as the continuous limit of a discrete model. Comparing the fixed shape g^* that we find in our analysis below with the diffusion matrix in (2.2.4), it turns out that the formulas coincide in $d = 1, 2$ but, remarkably, not in higher dimensions.

1. ('state space') $D \subset \mathbb{R}^d$ is a bounded open convex set, $\overline{D}$ is its closure and $\partial D = \overline{D} \backslash D$.

2. ('fixed shape') $g^* : \overline{D} \to \mathbb{R}$ is the unique continuous solution of

$$\begin{aligned} -\tfrac{1}{2}\Delta g^* &= 1 \qquad \text{on } D \\ g^* &= 0 \qquad \text{on } \partial D, \end{aligned} \tag{2.2.5}$$

   with $\Delta = \sum_{i=1}^{d} \frac{\partial^2}{\partial x_i{}^2}$ the Laplacian.

3. ('diffusion function') $\mathcal{H}$ is the class of functions $g : \overline{D} \to [0, \infty)$ satisfying

$$\begin{aligned} &\text{(i)} \quad g \leq M g^* \text{ for some } M < \infty \\ &\text{(ii)} \quad g > 0 \text{ on } D \\ &\text{(iii)} \quad g \text{ continuous on } \overline{D}. \end{aligned} \tag{2.2.6}$$

4. ('attraction point') $\theta \in \overline{D}$.

5. ('attraction constant') $c \in (0, \infty)$.

With these ingredients we let our basic diffusion equation be the SDE:

$$dX_t = c(\theta - X_t)dt + \sqrt{2g(X_t)}dB_t, \tag{2.2.7}$$

where $(B_t)_{t \geq 0}$ is standard d-dimensional Brownian motion. Solutions of (2.2.7) solve the martingale problem for the operator A with domain $\mathcal{D}(A)$ given by

$$\begin{aligned} (Af)(x) &:= \Big(c(\theta - x) \cdot \nabla + g(x)\Delta\Big) f(x) \\ \mathcal{D}(A) &:= \mathcal{C}^2(\overline{D}), \end{aligned} \tag{2.2.8}$$

where $\nabla = (\frac{\partial}{\partial x_i}, \ldots, \frac{\partial}{\partial x_d})$ and $\cdot$ denotes inner product. The martingale problem for A is well-posed if and only if, for each initial condition on $\overline{D}$, the SDE (2.2.7) has a unique weak solution $(X_t)_{t \geq 0}$. In this case, the operator A has a unique extension to a generator of a Feller semigroup, and $(X_t)_{t \geq 0}$ is the associated Feller process.[4]

By a continuous probability kernel on $\overline{D}$ we mean a continuous map $K : \overline{D} \to \mathcal{P}(\overline{D})$, written $x \mapsto K_x$, where $\mathcal{P}(\overline{D})$ is the space of probability measures on $\overline{D}$, equipped with the topology of weak convergence. We equip the space $\mathcal{K}(\overline{D}) := \mathcal{C}(\overline{D}, \mathcal{P}(\overline{D}))$ of probability kernels on $\overline{D}$ with the topology of uniform convergence. (Since $\mathcal{P}(\overline{D})$ is compact and Hausdorff, there is a unique uniform structure defining the topology, and we can unambiguously speak about uniform

[4] For a discussion of these facts, see the footnote at Theorem 2.1.1.

convergence of $\mathcal{P}(\overline{D})$-valued functions.) There exists a natural identification between continuous probability kernels $K \in \mathcal{K}(\overline{D})$ and continuous positive linear operators $K : \mathcal{C}(\overline{D}) \to \mathcal{C}(\overline{D})$ satisfying $K1 = 1$, the correspondence being given by

$$(Kf)(x) = \int_{\overline{D}} K_x(dy) f(y) \qquad (f \in \mathcal{C}(\overline{D})). \tag{2.2.9}$$

In this identification, the composition of two kernels is given by

$$(KL)_x(dy) = \int_{\overline{D}} K_x(dy) L_y(dz). \tag{2.2.10}$$

The convergence of operators $K_n \to K$ in the topology on $\mathcal{K}(\overline{D})$ is equivalent to the convergence of the functions $K_n f \to Kf$, uniformly on $\overline{D}$ for all $f \in \mathcal{C}(\overline{D})$.

In order to be able to define our renormalization transformation, we introduce a new class $\mathcal{H}'$ of diffusion functions as follows:

3′. $\mathcal{H}'$ is the class of all functions $g \in \mathcal{H}$ such that for all $c \in (0, \infty)$ and $\theta \in \overline{D}$: (1) The martingale problem associated with the operator A in (2.2.8) is well-posed. (2) The diffusion associated with (2.2.7) has a unique equilibrium $\nu_\theta^{g,c}$.

Here, by an equilibrium we mean a stationary distribution of (2.2.7). As we shall see in section 2.2.4, these assumptions are satisfied for many $g \in \mathcal{H}$. It turns out that the map $\theta \mapsto \nu_\theta^{g,c}$ is continuous, and so the equilibrium of (2.2.7) is a continuous probability kernel on $\overline{D}$ as a function of the parameter θ.

Theorem 2.2.1 *For each $g \in \mathcal{H}'$ and $c \in (0, \infty)$ there exists a continuous probability kernel $\nu^{g,c} \in \mathcal{K}(\overline{D})$ such that, for each $\theta \in \overline{D}$, $\nu_\theta^{g,c}$ is the equilibrium of the diffusion in (2.2.7).*

For $g \in \mathcal{H}'$ and $c \in (0, \infty)$, we now define our 'renormalization transformation' as

$$(F_c g)(\theta) := (\nu^{g,c} g)(\theta) = \int_{\overline{D}} g(x) \nu_\theta^{g,c}(dx). \tag{2.2.11}$$

In order to speak about the iterates of F_c, we need a subclass of $\mathcal{H}'$ that is closed under the F_c's. For this we may take the largest such subclass, so we define one more class of diffusion functions:

3″. $\mathcal{H}''$ is the union of all $\mathcal{G} \subset \mathcal{H}'$ such that $F_c(\mathcal{G}) \subset \mathcal{G}$ for all $c \in (0, \infty)$.

With these definitions, we have the following result.

Theorem 2.2.2 *For all $c \in (0, \infty)$: $F_c(\mathcal{H}') \subset \mathcal{H}$.*

It is at present not known if $\mathcal{H} = \mathcal{H}'$, but Theorem 2.2.2 implies at least that if (!) $\mathcal{H}' = \mathcal{H}$, then $\mathcal{H}'' = \mathcal{H}$.

The next result generalizes Theorem 2.1.7 (recall the composition of probability kernels defined in (2.2.10)):

Theorem 2.2.3 *For $g \in \mathcal{H}''$ and $k \geq 1$, let $K^{g,(k)}$ be given by*

$$K^{g,(k)} := \nu^{F^{(k-1)}g, c_k} \cdots \nu^{g,c_1}, \tag{2.2.12}$$

where $F^{(k)}g := (F_{c_k} \circ \cdots \circ F_{c_1})g$ is the k-th iterate of the renormalization transformations F_c applied to g ($F^{(0)}g = g$). If $\sum_k c_k^{-1} = \infty$, then in the sense of uniform convergence of probability kernels:

$$K^{g,(k)} \to K^{(\infty)} \quad \text{as } k \to \infty, \tag{2.2.13}$$

where the limiting kernel $K^{(\infty)}$ is universal in g and given by

$$K_\theta^{(\infty)}(dx) = P[B_\tau^\theta \in dx], \tag{2.2.14}$$

where $(B_t^\theta)_{t \geq 0}$ is Brownian motion starting in θ and $\tau := \inf\{t \geq 0 : B_t^\theta \in \partial D\}$.

The following generalizes Theorem 2.1.9:

Theorem 2.2.4
(a) *Let g^* be as in (2.2.5). If $g^* \in \mathcal{H}'$, then $rg^* \in \mathcal{H}''$ for all $r > 0$. Moreover, the 1-parameter family of functions rg^* ($r > 0$) are fixed shapes under F_c:*

$$F_c(rg^*) = \left(\frac{c}{c+r}\right) rg^*. \tag{2.2.15}$$

(b) *If $\sum_k c_k^{-1} = \infty$, then for all $g \in \mathcal{H}''$*

$$\lim_{k\to\infty} \sigma_k F^{(k)} g = g^* \quad \text{uniformly on } \overline{D}, \tag{2.2.16}$$

where $\sigma_k := \sum_{l=1}^k c_l^{-1}$.
(c) *If, in addition to the assumptions in (b), there exists a $\lambda > 0$ such that $g \geq \lambda g^*$, then*

$$\lim_{k\to\infty} \|\sigma_k F^{(k)} g - g^*\|_{\mathcal{H}} = 0, \tag{2.2.17}$$

where the norm $\|\cdot\|_{\mathcal{H}}$ is given by

$$\|g\|_{\mathcal{H}} := \sup_{x \in D} \left| \frac{g(x)}{g^*(x)} \right|. \tag{2.2.18}$$

In $d = 1$, formula (2.2.17) is in fact known to hold under somewhat weaker conditions on g (see Theorem 2.1.9 (c)).

2.2.4 Two renormalization classes: Theorems 2.2.5–2.2.10

The results in the last section are useful only after we come up with some examples of functions g in the classes $\mathcal{H}'$ and $\mathcal{H}''$. In this section we try to find sufficient conditions for $g \in \mathcal{H}'$ and for $g \in \mathcal{H}''$.

The following theorem shows that the assumption about ergodicity in the definition of $\mathcal{H}'$ is in most 'neat' cases satisfied.

Theorem 2.2.5 *Fix $g \in \mathcal{H}$, $\theta \in \overline{D}$ and $c \in (0, \infty)$. Assume that g is locally Hölder continuous (with positive exponent) on D, and that the martingale problem associated with the operator A in (2.2.8) is well-posed. Then the SDE (2.2.7) has a unique equilibrium $\nu_\theta^{g,c}$ and is ergodic, i.e., for each initial distribution the law of X_t converges weakly to $\nu_\theta^{g,c}$ as $t \to \infty$.*

Thus, for locally Hölder g, proving that $g \in \mathcal{H}'$ reduces to proving that the martingale problem for A in (2.2.8) is well-posed. As usual, existence of solutions is no problem:

Theorem 2.2.6 *For each $g \in \mathcal{H}$, $\theta \in \overline{D}$ and $c \in [0, \infty)$, and for each initial distribution on $\overline{D}$, the SDE (2.2.7) has a $\overline{D}$-valued, continuous, weak solution $(X_t)_{t \geq 0}$. The law of $(X_t)_{t \geq 0}$ solves the martingale problem for the operator A in (2.2.8).*

In fact, it seems reasonable to conjecture that uniqueness of the martingale problem, too, holds for all $g \in \mathcal{H}$, and (assuming ergodicity can also be proved), that $\mathcal{H}' = \mathcal{H}$. As we saw in Theorem 2.2.2, this would imply $\mathcal{H} = \mathcal{H}' = \mathcal{H}''$. However, it is not known whether uniqueness for the martingale problem holds for general $g \in \mathcal{H}$.[5]

In $d = 1$, uniqueness can be proved for many $g \in \mathcal{H}$, and the explicit formula for the equilibrium $\nu_\theta^{g,c}$ in (2.1.20) can be used to prove that, for all g, $F_c g$ is sufficiently nice. Indeed, the Yamada-Watanabe argument ([22], Proposition 5.2.13) and [1], Remark below Theorem 5, show that:

Theorem 2.2.7 *Assume that $d = 1$. If $g \in \mathcal{H}$ and $\sqrt{g}$ is Hölder $\frac{1}{2}$-continuous, then $g \in \mathcal{H}'$ and $g \in \mathcal{H}''$.*

In higher dimension, results are much harder to get. The standard theorem for strong uniqueness of multi-dimensional diffusions ([22], Theorem 5.2.9) and Theorem 2.2.5 give:

Theorem 2.2.8 *Assume that $d \geq 1$. If $g \in \mathcal{H}$ and $\sqrt{g}$ is Lipschitz, then $g \in \mathcal{H}'$.*

---

[5]Even in $d = 1$ this question seems to be open, although for each $g \in \mathcal{H}$ it is known that there exists a unique extension of A to a generator of a Feller semigroup. For this extended operator the martingale problem, of course, is well-posed.

If we restrict ourselves to initial conditions x and attraction points θ that lie within D, then the conditions for strong uniqueness can be weakened. We adopt the following definitions. If $x \in \partial D$, then $n(x) \in \mathbb{R}^d$ is called an (outward) normal to D in x if and only if $|n(x)| = 1$ and $\{y \in \mathbb{R}^d : (y - x) \cdot n(x) \geq 0\} \cap D = \emptyset$. A set $D \subset \mathbb{R}^d$ is called regular if and only if D is open, bounded, convex, and there exists a function $m \in \mathcal{C}^3(\overline{D})$, satisfying $m = 0$ on ∂D and $m < 0$ on D, with the property that for all $x \in \partial D$

$$|\nabla m(x)| = 1. \tag{2.2.19}$$

Note that, for each $x \in \partial D$, $\nabla m(x)$ is the unique normal to D in x. With these conventions we have the following theorem.

Theorem 2.2.9 *Let $g \in \mathcal{H}$, $\theta \in D$ and $c \in (0, \infty)$. Assume that D is a finite intersection of regular sets. Let g be Lipschitz on D, and assume that for all $x \in \partial D$, all $x_n \in D$ with $x_n \to x$, and each normal $n(x)$ to D in x:*

$$\limsup_{n\to\infty} \frac{g(x_n)}{|x - x_n|} < c(x - \theta) \cdot n(x). \tag{2.2.20}$$

Then any solution $(X(t))_{t\geq 0}$ of (2.2.7) with initial condition $X(0) = x$ ($x \in D$) satisfies

$$P[X(t) \in D \ \ \forall t \geq 0] = 1, \tag{2.2.21}$$

and strong uniqueness holds for the SDE (2.2.7) with initial condition x.

The idea behind this theorem is that, since $\sqrt{g}$ is locally Lipschitz on D, a modification of the standard proof for strong uniqueness shows that solutions of (2.2.7) are unique up to the first hitting time of the boundary, while condition (2.2.20) guarantees that this time is infinite. The essential difficulties in proving uniqueness occur when the diffusion hits the boundary in a finite time. Although the conditions on g in Theorem 2.2.9 are considerably weaker than those in Theorem 2.2.8, the result is still not very satisfactory for our purposes. Indeed, we want to vary θ, and so if (2.2.20) is to hold for all $\theta \in D$, then we must have 'sublinear' behavior of g at the boundary:

$$\lim_{n\to\infty} \frac{g(x_n)}{|x - x_n|} = 0 \qquad\qquad (x_n \in D,\ x_n \to x \in \partial D). \tag{2.2.22}$$

For example, it can be seen that for the fixed shape g^* condition (2.2.22) is violated. Consequently, Theorem 2.2.9 cannot even be used to give a satisfactory definition of $(F_c g^*)(\theta)$ for all $\theta \in D$.

Sufficient conditions for $g \in \mathcal{H}''$ are even harder to come by than sufficient conditions for $g \in \mathcal{H}'$. The following special case, however, shows us one example where $F_c g^*$ can be defined in a satisfactory way and where $\mathcal{H}''$ can be shown to be non-empty.

Theorem 2.2.10 *Let* $D = \{x \in \mathbb{R}^d : |x| < 1\}$. *Then* $g^*(x) = \frac{1}{d}(1 - |x|^2)$ *and* $g^* \in \mathcal{H}''$.

This last result is actually the only case in $d \geq 2$ where we are able to prove that $\mathcal{H}''$ is not empty. In view of Theorem 2.2.4 this is not a very satisfactory result, since we would like $\mathcal{H}''$ to at least contain a neighbourhood of g^* in order for the universality expressed in (2.2.16) to be meaningful. But nothing better is available at present.

2.2.5 Difficulties for $d \geq 2$

Higher-dimensional diffusions differ fundamentally from one-dimensional diffusions. In general they are technically much harder to treat. In our situation: let $(X_t^x)_{t\geq 0}$ and $(X_t^y)_{t\geq 0}$ be solutions of (2.2.7) with initial conditions x resp. y, adapted to the same Brownian motion, and let $g \in \mathcal{H}$ be Lipschitz. In $d = 1$ it can be shown (compare [9], equation (2.47)) that

$$E[|X_t^x - X_t^y|] \leq |x - y|e^{-ct}. \tag{2.2.23}$$

It is essentially with the help of this coupling that one is able to prove strong uniqueness for solutions of (2.2.7), convergence to equilibrium, and the property that the class $\mathcal{H}_{Lip}$ in (2.1.5) is closed under the transformation F_c. In $d \geq 2$, however, (2.2.23) does not hold. Indeed, let $(S_t)_{t\geq 0}$ be the semigroup associated with the process $(X_t)_{t\geq 0}$, i.e.

$$(S_t f)(x) = E[f(X_t^x)] \qquad (f \in \mathcal{C}(\overline{D})). \tag{2.2.24}$$

A direct consequence of (2.2.23) is the following: if f is Lipschitz with constant L, then $S_t f$ is Lipschitz with constant Le^{-ct}. However, in $d \geq 2$ it is possible to show that, for an appropriate g and c, there exist $t > 0$ and Lipschitz f such that the Lipschitz constant of $S_t f$ is strictly larger than the Lipschitz constant of f. Therefore (2.2.23) cannot hold for these g and c.

Thus, the diffusion (2.2.7) behaves differently in higher dimension in lacking a good coupling. It also differs in lacking reversibility. By definition, the diffusion in (2.2.7) is reversible if and only if its equilibrium $\nu_\theta^{g,c}(dx)$ solves the vector equation

$$\int_{\overline{D}} \nu_\theta^{g,c}(dx)[c(\theta - x) + g(x)\nabla]f(x) = 0 \qquad \forall f \in \mathcal{C}^1(\overline{D}). \tag{2.2.25}$$

Diffusions in $d = 1$ are typically reversible, and we can solve (2.2.25) explicitly for $\nu_\theta^{g,c}$ to get the formula (2.1.20). In $d \geq 2$, however, no matter what D, there exists no $g \in \mathcal{H}'$ such that (2.2.7) is reversible for all (!) $\theta \in \overline{D}$. Related to this is the fact that in general no explicit formula for $\nu_\theta^{g,c}$ is known. Similarly, for general D no explicit formulas are known for the limiting distribution $K_\theta^{(\infty)}$ and for the fixed shape g^*. Since for $d = 1$ the proofs of Theorems 2.1.7 and 2.1.9 were based on explicit manipulations with $\nu_\theta^{g,c}$ and g^* (see [1]), the generalization to $d \geq 2$ forces us to use more abstract methods in our proofs. We believe that these methods (in particular the proof of Lemma 2.3.4 and its use) also give a deeper understanding of the case $d = 1$.

2.2.6 Open problems

The most urgent open problems concern the question for which functions g it is possible to prove $g \in \mathcal{H}''$ (recall the definitions of $\mathcal{H}'$, $\mathcal{H}''$ and F_c in section 2.2.3). In particular, one may ask:

1. Is $g^* \in \mathcal{H}'$ for all bounded open convex D?

2. Is $\mathcal{H}' = \mathcal{H}$ for all bounded open convex D?

Since g^* is locally Hölder on D, it is sufficient for question 1. to show that uniqueness holds the martingale problem associated with A in (2.2.8) (by Theorems 2.2.5 and 2.2.6). If the answer to question 1. is affirmative, then at least $g^* \in \mathcal{H}''$ for all D (by Theorem 2.2.4 (a)). If the answer to question 2. is affirmative, then it implies that $\mathcal{H}'' = \mathcal{H}$, but question 2. certainly represents a hard problem.

In another approach, one may try to show that $\mathcal{H}''$ is not empty by deriving more properties for $F_c g$, given that g is nice. In analogy with the situation in $d = 1$, one may ask:

3. If $g \in \mathcal{H}$ is Lipschitz, then is $F_c g$ also Lipschitz?

4. If $g \in \mathcal{H}$ is Lipschitz, then does it follow that $g \in \mathcal{H}'$?

For question 3. one needs to control the behavior of the equilibrium $\nu_\theta^{g,c}$ as a function of θ. In the absense of an explicit formula, this can be attempted with coupling methods. In fact, the coupling that underlies Theorem 2.2.8 can be used to show that if $\sqrt{g}$ is Lipschitz with a sufficiently small Lipschitz constant, then $F_c g$ is Lipschitz. However, a better coupling than this one is hard to find in $d \geq 2$, and question 3. is still open. So is question 4., which is a well-known and hard open problem in the field.

2.3 The renormalization transformation

2.3.1 Notation

Let $E \subset \mathbb{R}^d$ be open or closed. By $B(E)$ we denote the bounded Borel-measurable real functions on E. For μ a finite measure on E and $f \in B(E)$ we write

$$\langle \mu | f \rangle := \int_E f d\mu. \tag{2.3.1}$$

The real continuous functions on E are denoted by $\mathcal{C}(E)$, and $\mathcal{C}_b(E)$ is the Banach space of bounded continuous functions with norm $\|f\| := \sup_{x\in E} |f(x)|$. By $\mathcal{C}^n(E)$ we denote the functions $f \in \mathcal{C}(E)$ such that all derivatives up to order n exist on the interior of E and can be extended to functions in $\mathcal{C}(E)$. By definition $\mathcal{C}^\infty(E) := \bigcap_n \mathcal{C}^n(E)$. We sometimes write $f \in \mathcal{C}^n(E)$ when we mean $f|_E \in \mathcal{C}^n(E)$, where $f|_E$ is the restriction of f to E. By $\mathcal{C}^n_c(E)$, $\mathcal{C}^\infty_c(E)$ we denote functions in $\mathcal{C}^n(E)$, $\mathcal{C}^\infty(E)$ that have a compact support in E.

When $X = (X_t)_{t\geq 0}$ is a continuous E-valued stochastic process and $\mathcal{F}^X_t := \sigma(X_s : s \in [0,t])$ is the filtration generated by X, we say that X solves the martingale problem for a linear operator A on $B(E)$ if and only if

$$f(X_t) - \int_0^t (Af)(X_s) ds \tag{2.3.2}$$

is an $\mathcal{F}^X_t$-martingale for all $f \in \mathcal{D}(A)$, the domain of A. We identify a linear operator A with domain $\mathcal{D}(A)$ with the linear space $\{(f, Af) : f \in \mathcal{D}(A)\}$. Closure always refers to the norm $\|f\|$. We say that a Feller semigroup $(S_t)_{t\geq 0}$ on $B(E)$ is related to X if and only if for all $f \in B(E)$ and $s, t \geq 0$

$$E[f(X_{t+s})|\mathcal{F}^X_t] = (S_t f)(X_s) \qquad a.s. \tag{2.3.3}$$

Finally, the notation $\mathcal{A}$ or $\mathcal{A}^{g,c}_\theta$ is used generally (without specification of the domain) for the differential form

$$(\mathcal{A}f)(x) := \Big(c(\theta - x)\cdot \nabla + g(x)\Delta\Big) f(x), \tag{2.3.4}$$

where $\nabla = (\frac{\partial}{\partial x_i}, \ldots, \frac{\partial}{\partial x_d})$, the $\cdot$ denotes inner product, and $\Delta = \nabla \cdot \nabla = \sum_{i=1}^d \frac{\partial^2}{\partial {x_i}^2}$ is the Laplacian. We write $|x| = \sqrt{x \cdot x}$ for the Euclidian norm.

2.3.2 Preliminaries

We begin with two lemmas collecting well-known facts. In section 2.2.3 we already mentioned the following.

Lemma 2.3.1 *Let $\mathcal{K}(\overline{D})$ be the set of continuous probability kernels on $\overline{D}$, equipped with the topology of uniform convergence, and let $\mathcal{K}'(\overline{D})$ be the space of all positive linear operators $K : \mathcal{C}(\overline{D}) \to \mathcal{C}(\overline{D})$ satisfying $K1 = 1$, equipped with the strong operator topology. Then a homeomorfism between $\mathcal{K}(\overline{D})$ and $\mathcal{K}'(\overline{D})$ is given by*

$$(Kf)(x) = \langle K_x | f \rangle \qquad (f \in \mathcal{C}(\overline{D})). \tag{2.3.5}$$

In particular, K_n converges to K in the topology on $\mathcal{K}(\overline{D})$ if and only if

$$\|K_n f - Kf\| \to 0 \quad \forall f \in \mathcal{C}(\overline{D}). \tag{2.3.6}$$

Proof of Lemma 2.3.1: Let $K \in \mathcal{K}(\overline{D})$, and for $f \in \mathcal{C}(\overline{D})$ define $(K'f)(x) := \langle K_x | f \rangle$. By the continuity of K, the function $x \mapsto \langle K_x | f \rangle$ is continuous. It is obvious that the map $f \mapsto K'f$ is positive and linear, and therefore continuous, and satisfies $K'1 = 1$. Conversely, by the Riesz-Markov theorem ([30], Theorem IV.14), each such K' defines a probability measure K_x for each $x \in \overline{D}$. Since $K'f \in \mathcal{C}(\overline{D})$ for each $f \in \mathcal{C}(\overline{D})$, the map $x \mapsto K_x$ is continuous in the weak topology. Finally, $\mathcal{P}(\overline{D})$ is continuously imbedded in $\mathcal{C}(\overline{D})^*$, the dual of $\mathcal{C}(\overline{D})$, so $\mathcal{K}(\overline{D})$ is continuously imbedded in $\mathcal{C}(\overline{D}, \mathcal{C}(\overline{D})^*)$, if we equip the latter with the topology of uniform convergence, where the uniform structure on $\mathcal{C}(\overline{D})^*$ is given by the semi-norms $l \mapsto |\langle l | f \rangle|$. The topology of uniform convergence in $\mathcal{C}(\overline{D}, \mathcal{C}(\overline{D})^*)$ is then defined by the semi-norms $(p_f)_{f \in \mathcal{C}(\overline{D})}$ given by

$$p_f(F) = \sup_{x \in \overline{D}} |\langle F(x) | f \rangle| \qquad (F \in \mathcal{C}(\overline{D}, \mathcal{C}(\overline{D})^*)). \tag{2.3.7}$$

If $K \in \mathcal{K}(\overline{D})$, then $p_f(K) = \sup_{x \in \overline{D}} |\langle K_x | f \rangle| = \|Kf\|$, so uniform convergence of probability kernels corresponds to convergence of the associated operators in the strong operator topology. From now on we identify kernels with linear operators as in (2.3.5).

Since $\overline{D}$ is convex, the Dirichlet problem on $\overline{D}$ always has a solution. We shall be interested in harmonic functions and functions of constant Laplacian.

Lemma 2.3.2 **(a)** *For every $\phi \in \mathcal{C}(\partial D)$ there exists a unique $f \in \mathcal{C}(\overline{D}) \cap \mathcal{C}^2(D)$ that solves*

$$\begin{array}{ll} \Delta f = 0 & \text{on } D \\ f = \phi & \text{on } \partial D. \end{array} \tag{2.3.8}$$

The solution is given by

$$f = H\phi, \tag{2.3.9}$$

where $H \in \mathcal{K}(\overline{D})$ is the probability kernel given by

$$H_x(dy) = P[B^x_\tau) \in dy], \tag{2.3.10}$$

where $(B_t^x)_{t\geq 0}$ *is Brownian motion starting at* x *and*

$$\tau := \inf\{t \geq 0 : B_t^x \in \partial D\}. \tag{2.3.11}$$

(b) *There exists a unique* $g^* \in \mathcal{C}(\overline{D}) \cap \mathcal{C}^2(D)$ *that solves*

$$\begin{aligned} -\tfrac{1}{2}\Delta g^* &= 1 \qquad \text{on } D \\ g^* &= 0 \qquad \text{on } \partial D. \end{aligned} \tag{2.3.12}$$

The solution is given (with τ *as in (2.3.11)) by*

$$g^*(x) = E^x[\tau] \tag{2.3.13}$$

and satisfies $g^* > 0$ *on* D*. There exists an* $L < \infty$ *such that*

$$g^*(x) \leq L|x - y| \qquad \forall x \in \overline{D},\ y \in \partial D. \tag{2.3.14}$$

Proof of Lemma 2.3.2: Formulas (2.3.9) and (2.3.10) can be found in [22], Proposition 4.2.7 and Theorems 4.2.12 and 4.2.19. For (2.3.13) see [22], Problem 4.2.25. The fact that $g^* > 0$ on D can easily be deduced from the representation (2.3.13), but alternatively one may consult [29], Theorem 2.5. To prove (2.3.14), we assume without loss of generality that $y = 0$ and $x_1 > 0$ $\forall x \in D$, where for any $x \in \mathbb{R}^d$ we write $x = (x_1, \ldots, x_d)$. Now choose L such that $|x - \tilde{x}| \leq L$ for all $x, \tilde{x} \in \overline{D}$. Define a stopping time $\tilde{\tau}$ by

$$\tilde{\tau} := \inf\{t \geq 0 : B_t^1 \in \{0, L\}\}, \tag{2.3.15}$$

where $B_t = (B_t^1, \ldots, B_t^d)$ is d-dimensional Brownian motion. By [22], Problem 4.2.25, we have

$$g^*(x) = E^x[\tau] \leq E^x[\tilde{\tau}] = x_1(L - x_1) \leq Lx_1 \leq L|x - y|. \tag{2.3.16}$$

∎

2.3.3 Proof of Theorem 2.2.1

Theorem 2.2.1 follows directly from the following lemma. Formula (2.3.17) (ii) below will be essential for the rest of this section.

Lemma 2.3.3 *Fix* $g \in \mathcal{H}'$ *and* $c \in (0, \infty)$*. For any* $\theta \in \overline{D}$*, denote by* $(S_t)_{t\geq 0}$ *the Feller semigroup related to the solution* $(X_t)_{t\geq 0}$ *of the martingale problem associated with* A *in (2.2.8), and let* G *be the full generator of* $(S_t)_{t\geq 0}$*. Then, for any*

$\theta \in \overline{D}$, the equilibrium $\nu_\theta^{g,c}$ of (2.2.7) is the unique solution of any of the following two equations:

$$\begin{array}{lll} \text{(i)} & \langle \nu_\theta^{g,c} | S_t f \rangle = \langle \nu_\theta^{g,c} | f \rangle & \forall t \geq 0,\ f \in \mathcal{C}(\overline{D}) \\ \text{(ii)} & \langle \nu_\theta^{g,c} | G f \rangle = 0 & \forall f \in \mathcal{D}(G).^6 \end{array} \tag{2.3.17}$$

For $\theta \in \partial D$, $\nu_\theta^{g,c} = \delta_\theta$ and for $\theta \in D$ the measure $\nu_\theta^{g,c}$ satisfies $\nu_\theta^{g,c}(D) > 0$. Furthermore, the map $\theta \mapsto \nu_\theta^{g,c}$ is continuous with respect to the topology of weak convergence.

Proof of Lemma 2.3.3: For simplicity we drop the superscripts g, c. Relation (2.3.17) (i) means that $E[f(X_t)]$ is independent of t when $(X_t)_{t\geq 0}$ is the solution of (2.2.7) with initial condition ν_θ. So (2.3.17) (i) just says that ν_θ is the unique equilibrium of (2.2.7), which is by definition true for $g \in \mathcal{H}'$. To prove (2.3.17) (ii), note that $Gf = \lim_{t\to 0} t^{-1}(S_t f - f)$ for all $f \in \mathcal{D}(G)$, where the limit is in the norm $\|\cdot\|$. So differentiating (2.3.17) (i), we get (2.3.17) (ii). To show that (2.3.17) (ii) determines ν_θ uniquely, note that for all $f \in \mathcal{D}(G)$ it holds that $S_t f \in \mathcal{D}(G)$ $\forall t \geq 0$ and $\frac{\partial}{\partial t} S_t f = G S_t f$, where the differentiation is in the Banach space $\mathcal{C}(\overline{D})$ (see [16], Proposition 1.1.5 (b)). Now, with $\tilde{\nu}_\theta$ a solution of (2.3.17) (ii), we have

$$\tfrac{\partial}{\partial t} \langle \tilde{\nu}_\theta | S_t f \rangle = \langle \tilde{\nu}_\theta | G S_t f \rangle = 0 \qquad \forall t \geq 0,\ \ f \in \mathcal{D}(G), \tag{2.3.18}$$

and this implies (2.3.17) (i) for $f \in \mathcal{D}(G)$. Since $\mathcal{D}(G)$ is dense in $\mathcal{C}(\overline{D})$, (2.3.17) (i) holds for general $f \in \mathcal{C}(\overline{D})$ and hence $\tilde{\nu}_\theta = \nu_\theta$.

To see that $\nu_\theta = \delta_\theta$ if $\theta \in \partial D$, note that $X_t \equiv \theta$ solves (2.2.7), so δ_θ is an equilibrium of (2.2.7). To see that $\nu_\theta(D) > 0$ for $\theta \in D$, insert $f(x) = |x - \theta|^2$ into (2.3.17) (ii) to get $c\langle \nu_\theta | f \rangle = d\langle \nu_\theta | g \rangle$ (compare also Lemma 2.3.4). Now f is strictly bounded away from zero on ∂D, so $\langle \nu_\theta | g \rangle > 0$. Since $g = 0$ on ∂D this implies $\nu_\theta(D) > 0$.

We next show that the probability kernel ν_θ is continuous in θ. For each $\theta \in \overline{D}$ let $(S_t^\theta)_{t\geq 0}$ be the Feller semigroup above and let G^θ be its generator. Let $\theta_n, \theta \in \overline{D}$ with $\theta_n \to \theta$. Using the fact that the martingale problem is well-posed for all θ, we have by [40], Theorem 11.1.4,

$$S_t^{\theta_n} f \to S_t^\theta f \qquad \forall f \in \mathcal{C}(\overline{D}),\ t \geq 0, \tag{2.3.19}$$

where the convergence is in $\mathcal{C}(\overline{D})$. By [16], Theorem 1.6.1 (c), it follows that for all $f \in \mathcal{D}(G^\theta)$ there exist $f_n \in \mathcal{D}(G^{\theta_n})$ such that

$$G^{\theta_n} f_n \to G^\theta f \quad \text{as } n \to \infty, \tag{2.3.20}$$

again in the topology on $\mathcal{C}(\overline{D})$. Now consider the sequence ν_{θ_n}. By compactness, it has a cluster point. For any such cluster point $\tilde{\nu}_\theta$, choose a subsequence such that ν_{θ_n} converges to $\tilde{\nu}_\theta$, and observe that for each $f \in \mathcal{D}(G^\theta)$, with f_n as in (2.3.20),

$$\begin{aligned}
&|\langle\tilde{\nu}_\theta|G^\theta f\rangle| \\
&\leq |\langle\tilde{\nu}_\theta|G^\theta f\rangle - \langle\nu_{\theta_n}|G^\theta f\rangle| + |\langle\nu_{\theta_n}|G^\theta f\rangle - \langle\nu_{\theta_n}|G^{\theta_n} f_n\rangle| + |\langle\nu_{\theta_n}|G^{\theta_n} f_n\rangle| \\
&\leq |\langle\tilde{\nu}_\theta|G^\theta f\rangle - \langle\nu_{\theta_n}|G^\theta f\rangle| + \|G^\theta f - G^{\theta_n} f_n\| + 0,
\end{aligned} \tag{2.3.21}$$

where the right-hand side tends to zero as $n \to \infty$. By (2.3.17) (ii), it follows that $\tilde{\nu}_\theta = \nu_\theta$ for each cluster point $\tilde{\nu}_\theta$ of the ν_{θ_n}, and hence ν_{θ_n} converges to ν_θ. ∎

2.3.4 Proof of Theorems 2.2.2–2.2.4

The proofs of Theorems 2.2.2–2.2.4 are based on the following lemma:

Lemma 2.3.4 *For any $g \in \mathcal{H}'$ and $c \in (0, \infty)$, let $\nu^{g,c} \in \mathcal{K}(\overline{D})$ as in Theorem 2.2.1. Fix $\lambda \in \mathbb{R}$. Assume that $f \in \mathcal{C}(\overline{D}) \cap \mathcal{C}^2(D)$ satisfies*

$$-\tfrac{1}{2}\Delta f = \lambda \qquad \text{on } D. \tag{2.3.22}$$

Then

$$\nu^{g,c} f = f - \tfrac{\lambda}{c}\,\nu^{g,c} g. \tag{2.3.23}$$

Proof of Lemma 2.3.4: We start with the case $f \in \mathcal{C}^2(\overline{D})$. Let $(T_t^{\theta,c})_{t\geq 0}$ be the Feller semigroup on $\mathcal{C}(\overline{D})$ defined by

$$(T^{\theta,c} f)(x) := f(\theta + e^{-ct}(x-\theta)) \qquad f \in \mathcal{C}(\overline{D}). \tag{2.3.24}$$

This is the semigroup related to our process in (2.2.7) when the local diffusion function g is set to zero. If $B_{\theta,c}$ is its full generator, then for every $f \in \mathcal{C}^1(\overline{D})$

$$(B_{\theta,c} f)(x) = c(\theta - x)\cdot\nabla f(x). \tag{2.3.25}$$

Let us introduce an operator that is in some sense an inverse to $B_{\theta,c}$. Define

$$\begin{aligned}
\mathcal{D}(B_{\theta,c}^{-1}) &:= \{f \in \mathcal{C}(\overline{D}) : \int_0^\infty \|T_t^{\theta,c} f\|\,dt < \infty\} \\
B_{\theta,c}^{-1} f &:= -\int_0^\infty T_t^{\theta,c} f\,dt.
\end{aligned} \tag{2.3.26}$$

It follows that

$$B_{\theta,c} B_{\theta,c}^{-1} f = f \qquad \forall f \in \mathcal{D}(B_{\theta,c}^{-1}), \tag{2.3.27}$$

as can be seen by writing (compare the proof of [16], Proposition 1.1.5 (a))

$$\begin{aligned} B_{\theta,c}B_{\theta,c}^{-1}f &= \lim_{\varepsilon\to 0} -\varepsilon^{-1}(T_\varepsilon^{\theta,c}-1)\int_0^\infty T_t^{\theta,c}f\,dt \\ &= \lim_{\varepsilon\to 0}\varepsilon^{-1}\int_0^\infty (T_t^{\theta,c}f - T_{t+\varepsilon}^{\theta,c})\,dt \\ &= \lim_{\varepsilon\to 0}\varepsilon^{-1}\left(\int_0^\infty T_t^{\theta,c}f\,dt - \int_\varepsilon^\infty T_t^{\theta,c}f\,dt\right) \\ &= \lim_{\varepsilon\to 0}\varepsilon^{-1}\int_0^\varepsilon T_t^{\theta,c}f\,dt = f. \end{aligned} \tag{2.3.28}$$

Now let $f \in \mathcal{C}^2(\overline{D})$, $-\frac{1}{2}\Delta f = \lambda$. Then

$$\begin{aligned} & f - f(\theta) \in \mathcal{D}(B_{\theta,c}^{-1}) \\ & B_{\theta,c}^{-1}(f - f(\theta)) \in \mathcal{C}^2(\overline{D}) \\ & \Delta B_{\theta,c}^{-1}(f - f(\theta)) = \tfrac{\lambda}{c}. \end{aligned} \tag{2.3.29}$$

To see this, substitute the variables $u = e^{-ct}$, $du = -ce^{-ct}dt$ into (2.3.26) to get

$$B_{\theta,c}^{-1}(f - f(\theta))(x) = -\int_0^1 \tfrac{1}{cu}\Big(f(\theta + u(x-\theta)) - f(\theta)\Big)du. \tag{2.3.30}$$

Since f is differentiable at θ, the integrand is bounded and it follows that $f - f(\theta) \in \mathcal{D}(B_{\theta,c}^{-1})$. Interchanging differentiation and integration, we get the following expressions for the derivatives of $B_{\theta,c}^{-1}(f - f(\theta))$:

$$\begin{aligned} \tfrac{\partial}{\partial x_i}B_{\theta,c}^{-1}f(x) &= -\int_0^1 \tfrac{1}{c}(\tfrac{\partial}{\partial x_i}f)(\theta + u(x-\theta))du \\ \tfrac{\partial^2}{\partial x_i\partial x_j}B_{\theta,c}^{-1}f(x) &= -\int_0^1 \tfrac{u}{c}(\tfrac{\partial^2}{\partial x_i\partial x_j}f)(\theta + u(x-\theta))du. \end{aligned} \tag{2.3.31}$$

The interchanging is allowed because the integrands on the right-hand sides are absolutely integrable. In particular, it follows that $\Delta B_{\theta,c}^{-1}(f - f(\theta)) = -\int_0^1 \frac{u}{c}\Delta(f - f(\theta))du = \int_0^1 \frac{u}{c}2\lambda du = \frac{\lambda}{c}$.

Applying (2.3.17) (ii) to the function $B_{\theta,c}^{-1}(f - f(\theta)) \in \mathcal{C}^2(\overline{D}) \subset \mathcal{D}(G)$, we get

$$0 = \langle \nu_\theta^{g,c}|(B_{\theta,c} + g\Delta)B_{\theta,c}^{-1}(f - f(\theta))\rangle = \langle \nu_\theta^{g,c}|f - f(\theta)\rangle + \langle \nu_\theta^{g,c}|\tfrac{\lambda}{c}g\rangle, \tag{2.3.32}$$

which gives (2.3.23). To extend formula (2.3.23) to $f \in \mathcal{C}(\overline{D}) \cap \mathcal{C}^2(D)$, pick an $x_0 \in D$ and a sequence $a_n \in (0,1)$ with $a_n \to 1$ as $n \to \infty$. Define functions $f_n \in \mathcal{C}^2(\overline{D})$ by

$$f_n(x) = a_n^{-2}f(x_0 + a_n(x - x_0)). \tag{2.3.33}$$

Then $-\frac{1}{2}\Delta f_n = \lambda$ for each n and $\|f_n - f\| \to 0$. Letting $n \to \infty$ and using the continuity of $\nu^{g,c}$, we conclude that (2.3.23) holds for f. ∎

We recall that by the definitions in section 2.2.3

$$\begin{aligned} &F_c g = \nu^{g,c} g \\ &F^{(k)} = F_{c_k} \circ \cdots \circ F_{c_1} \\ &K^{g,(k)} = \nu^{F^{(k-1)}g, c_k} \cdots \nu^{g,c_1}, \end{aligned} \tag{2.3.34}$$

so that

$$F^{(k)} g = K^{g,(k)} g. \tag{2.3.35}$$

The following lemma now follows easily by iterating Lemma 2.3.4.

Lemma 2.3.5 *Let $g \in \mathcal{H}'$, $c_1, \ldots, c_k \in (0, \infty)$, and let f be as in Lemma 2.3.4. Define $K^{g,(k)}$ and $F^{(k)}$ as in (2.3.34), and assume that $F^{(1)}g, \ldots, F^{(k-1)}g \in \mathcal{H}'$. Then*

$$K^{g,(k)} f = f - \lambda \sigma_k F^{(k)} g, \tag{2.3.36}$$

with $\sigma_k = \sum_{l=1}^{k} \frac{1}{c_l}$.

We are now ready to prove Theorems 2.2.2–2.2.4.

Proof of Theorem 2.2.2: Since $\nu^{g,c}$ is a continuous probability kernel, it follows from (2.3.34) that $F_c g \in \mathcal{C}(\overline{D})$. If $\theta \in \partial D$, then $\nu_\theta^{g,c} = \delta_\theta$ by Lemma 2.3.3 and so $(F_c g)(\theta) = 0$. If $\theta \in D$ then by the same lemma $\nu_\theta^{g,c}(D) > 0$ and so $(F_c g)(\theta) > 0$. Finally, inserting $f = g^*$ into Lemma 2.3.4, we get $F_c g = \nu^{g,c} g = cg^* - c\nu^{g,c} g^* \le cg^*$. ∎

Proof of Theorem 2.2.3: By Lemma 2.3.1, we must show that $K^{g,(k)} f \to Hf$ as $k \to \infty$ in the norm on $\mathcal{C}(\overline{D})$ for each $f \in \mathcal{C}(\overline{D})$, where H is defined by (2.3.10). By Lemma 2.3.5,

$$K^{g,(k)} g^* = g^* - \sigma_k K^{g,(k)} g. \tag{2.3.37}$$

It follows that $0 \le \sigma_k K^{g,(k)} g \le g^*$, and since $\sigma_k \to \infty$ we have $\|K^{g,(k)} g\| \to 0$. Since $g > 0$ on D, this in fact implies that for any $f \in \mathcal{C}(\overline{D})$ with $f \equiv 0$ on ∂D

$$\|K^{g,(k)} f\| \to 0. \tag{2.3.38}$$

To see why, define $R_n := \{x \in \overline{D} : \exists y \in \partial D,\ |x - y| < \frac{1}{n}\}$. Choose $\phi_n \in \mathcal{C}(\overline{D})$, $0 \le \phi_n \le 1$, such that $\phi_n \equiv 0$ on R_{n+1} and $\phi_n \equiv 1$ on $\overline{D} \backslash R_n$. For each n there exists an $M_n < \infty$ such that $\phi_n \le M_n g$, so $\|K^{g,(k)} \phi_n\| \to 0$ as $k \to \infty$. We may choose a subsequence $n_k \to \infty$ such that

$$\|K^{g,(k)} \phi_{n_k}\| \to 0 \qquad \text{as } k \to \infty. \tag{2.3.39}$$

Using this, we can estimate for f:

$$\|K^{g,(k)}f\| \leq \|f\| \cdot \|K^{g,(k)}\phi_{n_k}\| + \max_{x\in R_{n_k}} |f(x)|, \tag{2.3.40}$$

where the right-hand side tends to zero as $k \to \infty$. This proves (2.3.38).

For any $f \in \mathcal{C}(\overline{D})$ we can now write

$$K^{g,(k)}f = K^{g,(k)}(Hf - (Hf - f)) = Hf - K^{g,(k)}(Hf - f) \to Hf, \tag{2.3.41}$$

where we use (2.3.36) and (2.3.38). ∎

Proof of Theorem 2.2.4: Pick $g = f = rg^*$ in Lemma 2.3.4 to get

$$F_c(rg^*) = rg^* - \frac{r}{c}F_c(rg^*), \tag{2.3.42}$$

which implies Theorem 2.2.4 (a). To prove Theorem 2.2.4 (b) we observe that by Lemma 2.3.5,

$$\sigma_k F^{(k)}g = g^* - K^{g,(k)}g^*. \tag{2.3.43}$$

By (2.3.38), $\|K^{g,(k)}g^*\| \to 0$ as $k \to \infty$, and the theorem follows. To prove Theorem 2.2.4 (c), note that by the reasoning following (2.3.37),

$$\|K^{g,(k)}g\|_{\mathcal{H}} \to 0 \qquad \text{as } k \to \infty. \tag{2.3.44}$$

In the special case that $g \geq \lambda g^*$ for some $\lambda > 0$, it also follows that $\|K^{g,(k)}g^*\|_{\mathcal{H}} \to 0$ as $k \to \infty$. Inserting this into (2.3.43), we see that for such g, the convergence can be strengthened to

$$\|\sigma_k F^{(k)}g - g^*\|_{\mathcal{H}} \to 0 \qquad \text{as } k \to \infty. \tag{2.3.45}$$

∎

2.4 Ergodicity: Proof of Theorem 2.2.5

Theorem 2.2.5 follows from the following more technical lemma. In this section, we use the symbol ν for the probability measure $\nu_\theta^{g,c}$ (so ν denotes a probability measure, not a probability kernel).

Lemma 2.4.1 *Fix $g \in \mathcal{H}$, $\theta \in \overline{D}$ and $c \in (0,\infty)$. Assume that g is locally Hölder continuous (with positive exponent) on D and that the martingale problem* *associated with the operator A in (2.2.8) is well-posed. Then the SDE (2.2.7) has a unique equilibrium $\nu \in \mathcal{P}(\overline{D})$. Furthermore, for every $f \in \mathcal{C}(\overline{D})$*

$$\|S_t f - \langle \nu | f \rangle\| \to 0 \qquad \text{as } t \to \infty. \tag{2.4.1}$$

If $\theta \in D$, then there exist $t_0, r > 0$ such that (2.4.1) can be sharpened as follows: For all $f : \overline{D} \to [0,1]$ measurable

$$\|S_t f - \langle \nu | f \rangle\| \le e^{-r(t-t_0)}. \tag{2.4.2}$$

The proof of Lemma 2.4.1 is long and will keep us busy for the rest of this section. For notational simplicity we treat the case $c = 1$ only. Other c follow trivially by the scaling property $\mathcal{A}_\theta^{\lambda c, \lambda g} = \lambda \mathcal{A}_\theta^{c,g}$.

We start by proving (2.4.2). To that end we introduce two compact sets $\overline{B} \subset \overline{C} \subset D$, and prove that the expected time for X_t, starting from any point in $\overline{D}$, to reach into B is bounded uniformly in the starting point (Lemma 2.4.2). On C we then use results from the theory of non-degenerate diffusions to show that the distribution of the process starting from B can be bounded from below in a uniform way (Lemma 2.4.3). Combining these two results we arrive at Lemma 2.4.4, which shows that there exists a ν such that (2.4.2) holds. Once we have shown formula (2.4.2), it follows that (2.4.1) holds for $\theta \in D$. The case $\theta \in \partial D$ can then easily be treated separately. We end by showing that ν is the unique equilibrium of (2.2.7).

Without loss of generality we may assume $\theta = 0$. Choose $\varepsilon > 0$ such that $|x| \le 2\varepsilon \Rightarrow x \in D$ and define:

$$\begin{aligned} B &:= \{x \in \overline{D} : |x| < \varepsilon\} \\ C &:= \{x \in \overline{D} : |x| < 2\varepsilon\}. \end{aligned} \tag{2.4.3}$$

Lemma 2.4.2 *Let B be as in (2.4.3). Denote by $(X_t^x)_{t\ge 0}$ the process X starting at $X_0 = x$, and define a stopping time τ_B^x by*

$$\tau_B^x := \inf\{t \ge 0 : X_t \in \overline{B}\}. \tag{2.4.4}$$

Then there exists a constant $T < \infty$ such that

$$\sup_{x \in \overline{D}} E[\tau_B^x] \le T. \tag{2.4.5}$$

Proof of Lemma 2.4.2: Let h_d denote the function

$$h_d(x) := \begin{cases} -\log|x| & (d = 2) \\ (d-2)^{-1}|x|^{2-d} & (d \ne 2). \end{cases} \tag{2.4.6}$$

This function satisfies

$$\begin{aligned} \nabla h_d(x) &= -x|x|^{-d} \\ \Delta h_d(x) &= 0. \end{aligned} \tag{2.4.7}$$

For $\lambda \geq 0$, we define a function r_λ on $\overline{D}\backslash B$ by

$$r_\lambda(x) := -\log|x| + \lambda h_d(x). \tag{2.4.8}$$

We shall show that it is possible to choose λ such that $\mathcal{A}r_\lambda \geq 1$, with $\mathcal{A}$ the differential form in (2.3.4). Indeed, a little calculation shows that

$$\mathcal{A}r_\lambda(x) = 1 + \Big(\lambda + (2-d)g(x)|x|^{d-4}\Big)|x|^{2-d}. \tag{2.4.9}$$

and so we may choose

$$\begin{aligned} &\lambda = 0 && (d \leq 2) \\ &\lambda = \max\nolimits_{x\in\overline{D}\backslash B}\ g(x)|x|^{-1} && (d = 3) \\ &\lambda = \max\nolimits_{x\in\overline{D}}\ (d-2)g(x)|x|^{d-4} && (d \geq 4). \end{aligned} \tag{2.4.10}$$

Next, we can extend r_λ to a function in $\mathcal{C}^2(\overline{D})$, which now has the property (with A the operator in (2.2.8))

$$(Ar_\lambda)(x) \geq 1 \qquad (x \in \overline{D}\backslash B). \tag{2.4.11}$$

Abbreviate $\tau = \tau_B^x$ and let $r : [\varepsilon, \infty) \to \mathbb{R}$ be the (decreasing) function such that $r_\lambda(x) = r(|x|)$. The process X solves the martingale problem for A, so for each $x \in \overline{D}\backslash B$ and $t \geq 0$ we have

$$E[\tau \wedge t] \leq E\left[\int_0^{\tau\wedge t} (Ar_\lambda)(X_s)ds\right] = E[r(|X_{\tau\wedge t}|)] - r(|x|) \leq r(\varepsilon) - r(|x|). \tag{2.4.12}$$

The case $x \in B$ can be added trivially, and letting $t \uparrow \infty$ we find

$$E[\tau_B^x] \leq r(\varepsilon) - \min_{y\in\overline{D}} r(|y|) \qquad \forall x \in \overline{D}, \tag{2.4.13}$$

which completes the proof. ∎

We have shown that no matter where the process X starts in $\overline{D}$, it reaches into the set $\overline{B}$ in a finite expected time that is uniform in the starting point. We next turn our attention to the process starting in $\overline{B}$. We shall prove:

Lemma 2.4.3 *Let $(S_t)_{t\geq 0}$ be the Feller semigroup associated with X and let C be as in (2.4.3). For each $0 < t_1 < t_2$ there exists a non-zero finite measure μ on $\overline{D}$ such that*

$$\inf_{t\in[t_1,t_2]} \inf_{x\in\overline{B}} (S_t f)(x) \geq \langle f|\mu\rangle \qquad (f \geq 0,\ f \in \mathcal{C}(\overline{D})). \tag{2.4.14}$$

Proof of Lemma 2.4.3: We shall compare X with the process vanishing at ∂C. To that end, let

$$\tau := \inf\{t \geq 0 : X_t^x \in \overline{D} \backslash C\}. \tag{2.4.15}$$

Note that for any $f \in \mathcal{C}(\overline{D})$

$$(S_t f)(x) = E[f(X_t^x)] \geq E[f(X_t^x) 1_{\{t \leq \tau\}}]. \tag{2.4.16}$$

The function $(t, x) \mapsto E[f(X_t^x) 1_{\{t \leq \tau\}}]$ is the solution of a Cauchy problem on $[0, \infty) \times \overline{C}$ with Dirichlet boundary conditions on ∂C. Since the operator A is uniformly elliptic on $\overline{C}$ and the function g is Hölder continuous on $\overline{C}$, it is known (see [15], volume II, appendix §6, Theorem 0.6 and [19], Corollary 3.7.1) that a fundamental solution to this Cauchy problem exists. In particular, there exists a function $p \in \mathcal{C}((0, \infty) \times \overline{C} \times C)$ with the properties:

$$\begin{array}{ll} E[f(X_t^x) 1_{\{t \leq \tau\}}] = \int_C p_t(x|y) f(y) dy & (f \in \mathcal{C}(\overline{C})) \\ p_t(x|y) > 0 & ((t, x, y) \in (0, \infty) \times C \times C). \end{array} \tag{2.4.17}$$

Note that $p_t(x, \cdot)$ is the probability density of the process vanishing at ∂D. Applying (2.4.17), we get Lemma 2.4.3 if we choose for μ the measure on $\overline{C}$ given by

$$\begin{array}{l} \mu(dy) = \mu(y) dy \\ \mu(y) := \min\{p_t(x|y) : t \in [t_1, t_2],\ x \in \overline{B}\}. \end{array} \tag{2.4.18}$$

∎

Combining Lemmas 2.4.2 and 2.4.3 we get:

Lemma 2.4.4 *For all $\theta \in D$ there exists a $t_0 \in (0, \infty)$ and a non-zero finite measure μ on $\overline{D}$ such that, for all $f \in \mathcal{C}(\overline{D})$, $f \geq 0$,*

$$S_{t_0} f \geq \langle \mu | f \rangle. \tag{2.4.19}$$

Proof of Lemma 2.4.4: From Lemma 2.4.2 we get

$$P[\tau_B^x \leq 2T] \geq \tfrac{1}{2}. \tag{2.4.20}$$

Let $x \in \overline{D}$, and denote the distribution of $X_{\tau_B^x}^x$ by ρ. Let X_t^ρ be the process X with initial distribution ρ. By Lemma 2.4.3, there exists a μ such that

$$E[f(X_t^\rho)] \geq \langle f | \mu \rangle \qquad (t \in [T, 3T]). \tag{2.4.21}$$

By (2.4.20), (2.4.21) and the strong Markov property, we have for all $x \in \overline{D}$ and $f \in \mathcal{C}(\overline{D})$

$$\begin{aligned}
(S_{3T}f)(x) &\geq E[f(X_{3T}^x)1_{\{\tau_B^x \leq 2T\}}] \\
&= \int_0^{2T} E[f(X_{3T}^x)|\tau_B^x = s]P[\tau_B^x \in ds] \\
&= \int_0^{2T} E[f(X_{3T-s}^\rho)]P[\tau_B^x \in ds] \\
&\geq \langle f|\mu\rangle \int_0^{2T} P[\tau_B^x \in ds] \\
&\geq \tfrac{1}{2}\langle f|\mu\rangle,
\end{aligned} \tag{2.4.22}$$

and (2.4.19) follows if we replace μ by $\frac{1}{2}\mu$ and set $t_0 = 3T$. ∎

We have now completed the preparatory work and are ready for:

Proof of Lemma 2.4.1: From Lemma 2.4.4 we get (2.4.2) with a standard technique. This goes as follows. Fix a measurable $f : \overline{D} \to [0,1]$ and define, for $t \geq 0$,

$$\begin{aligned}
\nu_t^+ &:= \sup\nolimits_{x\in\overline{D}}(S_t f)(x) \\
\nu_t^- &:= \inf\nolimits_{x\in\overline{D}}(S_t f)(x).
\end{aligned} \tag{2.4.23}$$

By Lemma 2.4.4,

$$S_{t+t_0}f = S_{t_0}(\nu_t^+ - (\nu_t^+ - S_t f)) = \nu_t^+ - S_{t_0}(\nu_t^+ - S_t f) \leq \nu_t^+ - \langle\mu|\nu_t^+ - S_t f\rangle. \tag{2.4.24}$$

A similar argument applies to ν_t^-, and we get

$$\begin{aligned}
\nu_{t+t_0}^+ &\leq \nu_t^+ - \langle\mu|\nu_t^+ - S_t f\rangle \\
\nu_{t+t_0}^- &\geq \nu_t^- + \langle\mu|S_t f - \nu_t^-\rangle.
\end{aligned} \tag{2.4.25}$$

It follows that

$$\nu_{t+t_0}^+ - \nu_{t+t_0}^- \leq \nu_t^+ - \nu_t^- - \langle\mu|\nu_t^+ - \nu_t^-\rangle = (1 - \langle\mu|1\rangle)(\nu_t^+ - \nu_t^-), \tag{2.4.26}$$

and by induction that

$$\nu_{t+nt_0}^+ - \nu_{t+nt_0}^- \leq (1 - \langle\mu|1\rangle)^n. \tag{2.4.27}$$

We thus see that $S_t f$ converges uniformly to a constant. This constant we can formally denote by $\langle\nu|f\rangle$. Formula (2.4.2) now holds with $r = -t_0^{-1}\log(1 - \langle\mu|1\rangle)$.

To complete the proof of (2.4.2), it is left to show that $f \mapsto \langle\nu|f\rangle$, defined implicitly above for all measurable $f : \overline{D} \to [0,1]$, indeed corresponds to a probability measure. It is sufficient to show that the map is linear, positive, satisfies $\langle\nu|1\rangle = 1$, and is continuous with respect to increasing sequences of functions. The first three properties are easy. We therefore only show continuity.

Let $B_1(\overline{D}) := \{f : \overline{D} \to [0,1] : f \text{ measurable}\}$ and let $f_i, f_\infty \in B_1(\overline{D})$, $f_i \uparrow f_\infty$. Fix any probability measure ρ on $\overline{D}$. As $t \to \infty$, $S_t f$ converges at a rate that is uniform in $f \in B_1(\overline{D})$, so for every $\varepsilon > 0$ there exists a $t > 0$ such that

$$|\langle\rho|S_t f\rangle - \langle\nu|f\rangle| \le \varepsilon \quad \forall f \in B_1(\overline{D}). \tag{2.4.28}$$

There exists an n such that

$$|\langle\rho|S_t f_i\rangle - \langle\rho|S_t f_\infty\rangle| \le \varepsilon \quad \forall i \ge n, \tag{2.4.29}$$

and it follows that for every $i \ge n$

$$\begin{aligned}
&|\langle\nu|f_i\rangle - \langle\nu|f_\infty\rangle| \\
&\quad \le |\langle\nu|f_i\rangle - \langle\rho|S_t f_i\rangle| + |\langle\rho|S_t f_i\rangle - \langle\rho|S_t f_\infty\rangle| + |\langle\rho|S_t f_\infty\rangle - \langle\nu|f_\infty\rangle| \le 3\varepsilon.
\end{aligned} \tag{2.4.30}$$

Note that $\nu \ge \mu$, so $\nu(D) > 0$. This completes the proof of (2.4.2).

Trivially, (2.4.2) implies (2.4.1) for $\theta \in D$. We next turn to the proof of (2.4.1) for $\theta \in \partial D$. As we shall see, in this case ν turns out to be δ_θ. Let X_t be any solution to the martingale problem associated with A. For $x \in \overline{D}$ write $x = (x_1, \ldots, x_d)$, and write $X_t = (X_t^1, \ldots, X_t^d)$. Without loss of generality we may assume $\theta = 0$ and $x_1 > 0\ \forall x \in D$. From the martingale problem (2.3.2) we have for $i = 1, \ldots, d$

$$\begin{aligned}
E[X_t^i] &= E[X_0^i] - \textstyle\int_0^t E[X_s^i]ds \\
E[|X_t|^2] &= E[|X_0|^2] + 2d \textstyle\int_0^t E[g(X_s)]ds - 2\int_0^t E[|X_s|^2]ds.
\end{aligned} \tag{2.4.31}$$

We see immediately that

$$E[X_t^i] = E[X_0^i]e^{-t}. \tag{2.4.32}$$

By (2.2.6), and by (2.3.14) in Lemma 2.3.2 (b), there exists a constant $L < \infty$ such that

$$E[g(X_t)] \le L E[X_t^1]. \tag{2.4.33}$$

Let

$$M := \sup\{|x - y| : x, y \in \overline{D}\}. \tag{2.4.34}$$

Then (2.4.32) and (2.4.33) imply

$$E[g(X_t)] \le LMe^{-t}. \tag{2.4.35}$$

Next, the function $t \mapsto E[|X_t|^2]$ is differentiable and satisfies

$$\tfrac{\partial}{\partial t}E[|X_t|^2] = 2dE[g(X_t)] - 2E[|X_t|^2]. \tag{2.4.36}$$

From this it follows that

$$\frac{\partial}{\partial t}\Big(E[|X_t|^2]e^{2t}\Big) = 2dE[g(X_t)]e^{2t} \tag{2.4.37}$$

and therefore

$$E[|X_t|^2] \le e^{-2t}E[|X_0|^2] + e^{-2t}\int_0^\infty 2dE[g(X_s)]e^{2s}ds. \tag{2.4.38}$$

Here, by (2.4.35),

$$\begin{aligned} &e^{-2t}\int_0^\infty 2dE[g(X_s)]e^{2s}ds = \int_0^\infty 2dE[g(X_s)]e^{2(t-s)}ds \\ &\le e^{-t}\int_0^t 2dE[g(X_s)]ds + \int_t^\infty 2dE[g(X_s)]ds \le 2dLMe^{-t} + 2dLMe^{-t}. \end{aligned} \tag{2.4.39}$$

Hence, combining (2.4.38) and (2.4.39) we get

$$E[|X_t|^2] \le e^{-2t}M^2 + 4dLMe^{-t}, \tag{2.4.40}$$

which tends to zero as $t \to \infty$. Since also $E[X_t^i] \to 0$ as $t \to \infty$, Chebyshev's inequality shows that $(S_t f)(x) = E[f(X_t^x)]$ converges to $\langle \delta_0 | f\rangle = f(0)$ for each $f \in \mathcal{C}(\overline{D})$, and (2.4.40) shows that this convergence is uniform in the initial value x. This completes the proof of (2.4.1).

We complete the proof of the theorem by showing that ν is the unique equilibrium of (2.2.7). This means that we must show (compare (2.3.17) (i)) that ν is the unique solution of

$$\langle \nu | S_t f\rangle = \langle \nu | f\rangle \qquad \forall t \ge 0,\ f \in \mathcal{C}(\overline{D}). \tag{2.4.41}$$

First, for any $x \in \overline{D}$,

$$\langle \nu | S_t f\rangle = \lim_{s\to\infty}(S_s S_t f)(x) = \lim_{s\to\infty}(S_{s+t} f)(x) = \langle \nu | f\rangle, \tag{2.4.42}$$

which proves that (2.4.41) holds. Suppose that $\tilde\nu$ is another solution. Let $t \to \infty$ in (2.4.41) and use that $S_t f \to \langle \nu | f\rangle$. By dominated convergence, $\langle \tilde\nu | S_t f\rangle \to \langle \nu | f\rangle$. So $\langle \tilde\nu | f\rangle = \langle \nu | f\rangle$ for all $f \in \mathcal{C}(\overline{D})$, and hence $\tilde\nu = \nu$. ∎

Remark Formula (2.4.2) actually shows that $\nu(D) = 1$ for $\theta \in D$, whenever it is true that for $x \in D$

$$P[X_t^x \in D] = 1 \qquad \forall t \ge 0. \tag{2.4.43}$$

Formula (2.4.43) holds, for example, under the conditions of Theorem 2.2.9, but no doubt much more generally too.

2.5 The martingale problem

In this section we prove the theorems about the martingale problem for A mentioned in section 2.2.4. The proofs of Theorems 2.2.7 and 2.2.8 have already been indicated in the text.

2.5.1 Existence: Proof of Theorem 2.2.6

We extend the function g to $\mathbb{R}^d$ by putting $g \equiv 0$ on $\mathbb{R}^d \backslash \overline{D}$. Let $\mu \in \mathcal{P}(\overline{D})$. By [16], Theorem 5.3.10, there exists an $\mathbb{R}^d$-valued weak solution to the SDE

$$dX_t = c(\theta - X_t)dt + \sqrt{2g(X_t)}dB_t \tag{2.5.1}$$

with initial distribution $P[X_0 \in dx] = \mu(dx)$. By the same theorem, X solves the martingale problem for the operator $\{(f, \mathcal{A}f) : f \in \mathcal{C}_c^\infty(\mathbb{R}^d)\}$. By [16], Proposition 7.1 from the appendix, there exist $f_n \in \mathcal{C}_c^\infty(\mathbb{R}^d)$ such that $f_n \to f$ and $\mathcal{A}f_n \to \mathcal{A}f$ uniformly on $\mathbb{R}^d$. By [16], Lemma 4.5.1, X now also solves the martingale problem for $\{(f, \mathcal{A}f) : f \in \mathcal{C}_c^2(\mathbb{R}^d)\}$.

Pick $x_i \in \overline{D}$, $R_i \in (0, \infty)$ such that $\overline{D} = \bigcap_i \{x \in \mathbb{R}^d : |x - x_i| \leq R_i\}$. Let $h \in \mathcal{C}^2(\mathbb{R})$, $h \equiv 1$ on $(-\infty, 0]$, $h \equiv 0$ on $[1, \infty)$ and $h' \leq 0$. Define $f_i \in \mathcal{C}_c^2(\mathbb{R}^d)$ by $f_i(x) := h(|x - x_i| - R_i)$. Then $f_i \in \mathcal{C}_c^2(\mathbb{R}^d)$ and it is easy to see that $\mathcal{A}f_i \geq 0$. By the martingale problem,

$$E[f_i(X_t)] = 1 + E[\int_0^t (\mathcal{A}f_i)(X_s)ds] \geq 1, \tag{2.5.2}$$

which shows that $P[|X_t - x_i| \leq R_i] = 1 \ \forall t \geq 0 \ \forall i$. By the continuity of X, it follows that $P[|X_t - x_i| \leq R_i \ \forall t \geq 0, i] = 1$ and therefore

$$P[X_t \in \overline{D} \ \forall t \geq 0] = 1. \tag{2.5.3}$$

By Whitney's extension theorem ([16], Corollary 6.3 in the appendix) it now follows that $\mathcal{C}^2(\overline{D}) = \{f|_{\overline{D}} : f \in \mathcal{C}_c^2(\mathbb{R}^d)\}$, and therefore X solves the martingale problem for $A = \{(f, \mathcal{A}f) : f \in \mathcal{C}^2(\overline{D})\}$. ∎

2.5.2 Strong uniqueness: Proof of Theorem 2.2.9

For notational simplicity we only consider the case $c = 1$ and $\theta = 0$. Our first aim is to prove (2.2.21), i.e., we show that the time needed for X_t to reach the boundary ∂D is infinite (Lemma 2.5.5). For this we construct (in Lemmas 2.5.3 and 2.5.4) a function h on D such that $\mathcal{A}h \leq 1$, where $\mathcal{A}$ is the differential form in (2.3.4), i.e.,

$$(\mathcal{A}f)(x) = \Big(- x \cdot \nabla + g(x)\Delta \Big) f(x). \tag{2.5.4}$$

With the help of a radial function (Lemma 2.5.2) the problem is reduced to a one-dimensional problem (Lemma 2.5.1).

Lemma 2.5.1 *Let $a, b \in \mathcal{C}[0,1]$ and $a > 0$ on $(0,1]$. Then there exists a unique function $f \in \mathcal{C}^2(0,1]$ such that*

$$\begin{aligned} &f(1) = f'(1) = 0 \\ &b(r)f'(r) + a(r)f''(r) = 1. \end{aligned} \tag{2.5.5}$$

For all $r \in (0,1)$ this function satisfies

$$\begin{aligned} &f(r) > 0 \\ &f'(r) < 0. \end{aligned} \tag{2.5.6}$$

Furthermore, if

$$\limsup_{r \to 0} \frac{a(r)}{r} < b(0), \tag{2.5.7}$$

then

$$\lim_{r \to 0} f(r) = \infty. \tag{2.5.8}$$

Proof of Lemma 2.5.1: Let $u \in \mathcal{C}^2(0,1]$ be the unique solution of

$$\begin{aligned} &u(1) = 0 \\ &u'(1) = -1 \\ &b(r)u'(r) + a(r)u''(r) = 0, \end{aligned} \tag{2.5.9}$$

i.e.,

$$\begin{aligned} u(r) &= -\textstyle\int_1^r dx \exp(-\int_1^x dy \frac{b(y)}{a(y)}) \\ u'(r) &= -\exp(-\textstyle\int_1^r dx \frac{b(x)}{a(x)}) \\ u''(r) &= \textstyle\frac{b(r)}{a(r)} \exp(-\int_1^r dx \frac{b(x)}{a(x)}). \end{aligned} \tag{2.5.10}$$

Note that $u(r) \geq 0$ and $u'(r) < 0$ for all $r \in (0,1]$. From the latter property it follows that u is invertible. Let $u(0) := \lim_{r \to 0} u(r)$ (which is allowed to be ∞). There exists a continuous function $v : [0, u(0)) \to (0, \infty)$ such that

$$v(u(r)) = a(r)(u'(r))^2. \tag{2.5.11}$$

Let $h \in \mathcal{C}[0, u(0))$ be the unique solution of

$$\begin{aligned} &h(0) = h'(0) = 0 \\ &v(u)h''(u) = 1 \qquad (u \in [0, u(0))), \end{aligned} \tag{2.5.12}$$

i.e.,

$$h(u) = \int_0^u dp \int_0^p dq \, \frac{1}{v(q)}. \tag{2.5.13}$$

Note that $h(u) > 0$ and $h'(u) > 0$ for all $u \in (0, u(0))$. We now define $f \in \mathcal{C}^2(0, 1]$ by

$$f(r) := h(u(r)). \tag{2.5.14}$$

It follows that

$$\begin{aligned} \frac{\partial}{\partial r} h(u(r)) &= h'(u(r))u'(r) \\ \frac{\partial^2}{\partial r^2} h(u(r)) &= h''(u(r))(u'(r))^2 + h'(u(r))u''(r) \\ b(r)f'(r) + a(r)f''(r) &= \Big(b(r)u'(r) + a(r)u''(r)\Big)h'(u(r)) \\ &\quad + a(r)(u'(r))^2 h''(u(r)) \\ &= v(u(r))h''(u(r)) = 1. \end{aligned} \tag{2.5.15}$$

We see that f constructed above is the unique solution of (2.5.5), and that f satisfies (2.5.6).

It is left to show that, under the conditions mentioned, f diverges as $r \to 0$. Let L be such that $\limsup_{r\to 0} r^{-1}a(r) < L < b(0)$. It follows that there exists an $\varepsilon > 0$ such that $b(r) > L$ and $a(r) < Lr$ for all $r \in [0, \varepsilon]$. Let $\tilde{f} \in \mathcal{C}(0, \varepsilon]$ be the unique solution of

$$\begin{aligned} &\tilde{f}(\varepsilon) = 0 \\ &\tilde{f}'(\varepsilon) = 0 \\ &L\tilde{f}'(r) + Lr\tilde{f}''(r) = 1, \end{aligned} \tag{2.5.16}$$

i.e.,

$$\begin{aligned} &(Lr\tilde{f}'(r))' = 1 \\ &Lr\tilde{f}'(r) = r - \varepsilon \\ &\tilde{f}'(r) = \tfrac{1}{L}(1 - \tfrac{\varepsilon}{r}) \\ &\tilde{f}(r) = \tfrac{1}{L}r - \varepsilon\log(r) - \varepsilon + \varepsilon\log(\varepsilon). \end{aligned} \tag{2.5.17}$$

It is clear that $\tilde{f}(r) \to \infty$ as $r \to 0$. Furthermore,

$$b(r)\tilde{f}'(r) + a(r)\tilde{f}''(r) < 1. \tag{2.5.18}$$

On $(0, \varepsilon]$ define $h := f - \tilde{f}$. Then, using (2.5.6), we get

$$\begin{aligned} &h(\varepsilon) > 0 \\ &h'(\varepsilon) < 0 \\ &b(r)h'(r) + a(r)h''(r) > 0. \end{aligned} \tag{2.5.19}$$

It follows that $h(r) > 0$ for all $r \in (0, \varepsilon]$: if we assume the converse, then h must assume a positive maximum in a point $0 < r < \varepsilon$, which is impossible by (2.5.19). We thus see that $f > \tilde{f}$, and therefore $f(r) \to \infty$ if $r \to 0$. ∎

Lemma 2.5.2 *Let D be regular and $0 \in D$. Then there exist a function $r \in \mathcal{C}^2(\overline{D})$ and a constant $K \in (0, \infty)$ with the following properties:*

$$\begin{array}{ll} 0 < r(x) \leq 1 & (x \in D) \\ r(x) = 0 & (x \in \partial D) \\ -x \cdot \nabla r(x) = K & (x \in \partial D). \end{array} \tag{2.5.20}$$

Proof of Lemma 2.5.2: Recall the definition of a regular set in section 2.2.4 and the function m associated with it. The function $x \mapsto x \cdot n(x) = x \cdot \nabla m(x)$ is $\mathcal{C}^2$ and strictly positive on ∂D, so we can find a strictly positive function $\phi \in \mathcal{C}^2(\overline{D})$ such that in an open neighbourhood of ∂D:

$$\phi(x) x \cdot n(x) = 1. \tag{2.5.21}$$

Define

$$r(x) := -\phi(x) m(x) \qquad (x \in D). \tag{2.5.22}$$

Then $r \in \mathcal{C}^2(\overline{D})$ and, for all $x \in \partial D$, $\nabla r(x)$ is parallel to $n(x)$ and satisfies $-x \cdot \nabla r(x) = \phi(x) x \cdot \nabla m(x) = 1$. We can multiply r with a constant to get $r \leq 1$. ∎

Lemma 2.5.3 *Let $D' \supset D$ be regular. For $x \in \partial D'$, let $n(x)$ be the normal to D' in x. For $x \in \overline{D}$, let*

$$l(x) := \inf\{|x - y| : y \in \partial D\}. \tag{2.5.23}$$

Assume that, for all $x_n \in D$, $x \in \partial D' \cap \partial D$ with $x_n \to x$ as $n \to \infty$,

$$\limsup_{n \to \infty} \frac{g(x_n)}{l(x_n)} < x \cdot n(x). \tag{2.5.24}$$

Then there exists a function $h \in \mathcal{C}(D')$ such that

$$\begin{array}{ll} 0 \leq h(x) & (x \in D') \\ \lim_{n \to \infty} h(x_n) = \infty & (x_n \to x \in \partial D') \\ (\mathcal{A}h)(x) \leq 1 & (x \in D). \end{array} \tag{2.5.25}$$

Proof of Lemma 2.5.3: Extend g by putting $g \equiv 0$ on $D' \backslash D$, so that (2.5.24) holds for all $x_n \in D'$ with $x_n \to x \in \partial D'$. Let r be as in Lemma 2.5.2. The idea will be to find a function $f \in \mathcal{C}^2(0, 1]$ such that

$$h(x) := f(r(x)) \tag{2.5.26}$$

satisfies (2.5.25).

For any $f \in \mathcal{C}^2(0,1]$ we have

$$\begin{aligned}
\nabla f(r(x)) &= f'(r(x))(\nabla r)(x)\\
\Delta f(r(x)) &= f''(r(x))(\nabla r)(x)\cdot(\nabla r)(x) + f'(r(x))(\Delta r)(x)\\
\mathcal{A}f(r(x)) &= \Big(-x\cdot(\nabla r)(x) + g(x)(\Delta r)(x)\Big)f'(r(x))\\
&\quad +\Big(g(x)|\nabla r)(x)|^2\Big)f''(r(x)),
\end{aligned} \tag{2.5.27}$$

where the first two formulas follow from

$$\begin{aligned}
&\tfrac{\partial}{\partial x_i}f(r(x)) = f'(r(x))\tfrac{\partial}{\partial x_i}r(x)\\
&\textstyle\sum_i \tfrac{\partial^2}{\partial x_i{}^2}f(r(x)) = \sum_i \tfrac{\partial}{\partial x_i}f'(r(x))\tfrac{\partial}{\partial x_i}r(x)\\
&= \textstyle\sum_i (f''(r(x))(\tfrac{\partial}{\partial x_i}r(x))(\tfrac{\partial}{\partial x_i}r(x)) + f'(r(x))\tfrac{\partial^2}{\partial x_i{}^2}r(x)).
\end{aligned} \tag{2.5.28}$$

We want estimates on the two terms in the formula for $\mathcal{A}f(r(x))$. To that aim, we define functions $a, b \in \mathcal{C}[0,1]$ by

$$\begin{aligned}
a(z) &:= \max\{g(x)|\nabla r(x)|^2 : r(x) = z\}\\
b(z) &:= \min\{-x\cdot\nabla r(x) + g(x)\Delta r(x) : r(x) = z\},
\end{aligned} \tag{2.5.29}$$

We have

$$\begin{aligned}
&b(0) = K\\
&\limsup_{z\to 0}\frac{a(z)}{z} < K.
\end{aligned} \tag{2.5.30}$$

Indeed, the first equation is trivial. For the second one, note that

$$\frac{a(z)}{z} = \max\Big\{\frac{g(x)|\nabla r(x)|^2}{r(x)} : r(x) = z\Big\}, \tag{2.5.31}$$

where, by (2.5.24), for each $x_n \in D'$ with $x_n \to x \in \partial D'$

$$\begin{aligned}
\limsup_{n\to\infty}\frac{g(x_n)|\nabla r(x_n)|^2}{r(x_n)} &= \Big(\limsup_{n\to\infty}\frac{g(x_n)}{l(x_n)}\Big)\Big(\lim_{n\to\infty}\frac{l(x_n)}{r(x_n)}|\nabla r(x_n)|^2\Big)\\
&< (x\cdot n(x))|\nabla r(x)| = -x\cdot\nabla r(x) = K.
\end{aligned} \tag{2.5.32}$$

Here, the last two equalities follow from Lemma 2.5.2. Using compactness, we arrive at (2.5.30).

We have thus found functions $a, b \in \mathcal{C}[0,1]$ such that

$$\begin{aligned}
-x\cdot\nabla r(x) + g(x)\Delta r(x) &\geq b(r(x))\\
g(x)|\nabla r(x)|^2 &\leq a(r(x))\\
\limsup_{z\to 0}\tfrac{a(z)}{z} &< b(0).
\end{aligned} \tag{2.5.33}$$

We can change a such that $a > 0$ on $(0, 1]$ while (2.5.33) continues to holds. Applying Lemma 2.5.1 to these functions a and b, we find a function f satisfying (2.5.5), (2.5.6) and (2.5.8). Since $b(0) > 0$, we see that there exists an $\varepsilon > 0$ such that

$$f''(z) > 0 \qquad (z \in (0, \varepsilon)). \tag{2.5.34}$$

Combining this with (2.5.6), (2.5.27) and (2.5.33), we see that

$$(\mathcal{A}h)(x) \leq 1 \qquad (x \in D, r(x) < \varepsilon), \tag{2.5.35}$$

where $h(x) := f(r(x))$ as in (2.5.26). But $x \mapsto (\mathcal{A}h)(x)$ is continuous on the compact set $\{x \in D : r(x) \geq \varepsilon\}$, so multiplying h by a constant we arrive at a function satisfying (2.5.25). ∎

Lemma 2.5.4 *Let D be a finite intersection of regular sets. Assume that for all $x \in \partial D$, all $x_n \in D$ with $x_n \to x$, and each normal $n(x)$ to D in x:*

$$\limsup_{n\to\infty} \frac{g(x_n)}{|x - x_n|} < x \cdot n(x). \tag{2.5.36}$$

Then there exists a function $h \in \mathcal{C}^2(D)$ such that

$$\begin{aligned} 0 &\leq h(x) \\ (\mathcal{A}h)(x) &\leq 1 \end{aligned} \tag{2.5.37}$$

and such that $h(x_n) \to \infty$ for all $x_n \to x \in \partial D$.

Proof of Lemma 2.5.4: Let $D = \bigcap_{i=1}^n D_i$, where the D_i are regular. For each D_i the assumptions in Lemma 2.5.3 are satisfied. In particular, (2.5.36) implies (2.5.24). Let $h_i \in \mathcal{C}^2(D_i)$ be the function constructed in Lemma 2.5.3. Then $h = \frac{1}{n}\sum_{i=1}^n h_i$ satisfies our requirements. ∎

Lemma 2.5.5 *Let D and g be as in Lemma 2.5.4, and let $(X^x_t)_{t\geq 0}$ be a solution to the martingale problem for A with $X^x_0 = x \in D$. Then*

$$P[X^x_t \in D \;\; \forall t \geq 0] = 1. \tag{2.5.38}$$

Proof of Lemma 2.5.5: Let h be the function mentioned in Lemma 2.5.3. For $H < \infty$ we introduce a stopping time τ_H by

$$\tau_H := \inf\{t \geq 0 : h(X^x_t) = H\}. \tag{2.5.39}$$

We can extend h outside $\{x \in \overline{D} : h(x) \leq H\}$ to a function in $\mathcal{C}^2(\overline{D})$. From the martingale problem we get

$$E[h(X_{t\wedge\tau_H})] = h(x) + E[\int_0^{t\wedge\tau_H} (\mathcal{A}h)(X_s)ds] \leq h(x) + E[t \wedge \tau_H]. \quad (2.5.40)$$

Here $E[h(X_{t\wedge\tau_H})] \geq HP[\tau_H \leq t]$ and $E[t \wedge \tau_H] \leq t$, so

$$P[\tau_H \leq t] \leq \frac{h(x)+t}{H}. \quad (2.5.41)$$

Therefore

$$P[h(X_s^x) < H \ \ \forall 0 \leq s \leq t] \geq 1 - \frac{h(x)+t}{H}. \quad (2.5.42)$$

Letting $H \uparrow \infty$ so that $\{x \in \overline{D} : h(x) < H\} \uparrow D$, we find that

$$P[X_s^x \in D \ \ \forall 0 \leq s \leq t] = 1. \quad (2.5.43)$$

Letting $t \uparrow \infty$ we obtain Lemma 2.5.5. ∎

Proof of Theorem 2.2.9: Define $D_n := \{x \in D : g(x) > \frac{1}{n}\}$. Note that if g is Lipschitz on $\overline{D}_n$ with constant L_n, then $\sqrt{g}$ is Lipschitz on $\overline{D}_n$ with constant nL_n. Let $(X_t)_{t\geq 0}$ and $(\tilde{X}_t)_{t\geq 0}$ be solutions of (2.2.7) with $X_0 = \tilde{X}_0$, adapted to the same Brownian motion. Define stopping times

$$\tau_n := \inf\{t \geq 0 : X_t \in \partial D_n\}, \quad (2.5.44)$$

and define $\tilde{\tau}_n$ similarly for $(\tilde{X}_t)_{t\geq 0}$. Now follow the proof of Theorem 5.2.5 in [22], to see that the processes X and $\tilde{X}$ are indistinguishable up to time $\tau_n \wedge \tau_n$, where by Lemma 2.5.5, $\tau_n \wedge \tilde{\tau}_n \uparrow \infty$ as $n \uparrow \infty$. ∎

2.5.3 Weak uniqueness: Proof of Theorem 2.2.10

Writing

$$\begin{aligned} -\tfrac{1}{2}\Delta\tfrac{1}{d}(1-|x|^2) &= -\tfrac{1}{2}\tfrac{1}{d}\textstyle\sum_{i=1}^d \frac{\partial^2}{\partial {x_i}^2}(1-\sum_{j=1}^d x_j^2) \\ &= \tfrac{1}{2d}\textstyle\sum_{i=1}^d \frac{\partial}{\partial x_i}(2\sum_{j=1}^d \delta_{ij}x_j) = \frac{1}{2d}\sum_{i=1}^d 2 = 1, \end{aligned} \quad (2.5.45)$$

we see that $g^*(x) = \frac{1}{d}(1-|x|^2)$ as claimed. We introduce polynomials on $\overline{D}$ in the usual way, namely we define the set of all multi-indices α by

$$\begin{aligned} A &:= \{\alpha \in \mathbb{Z}^d : \alpha_i \geq 0 \ \forall i = 1,\ldots,d\} \\ |\alpha| &:= \textstyle\sum_{i=1}^d \alpha_i. \end{aligned} \quad (2.5.46)$$

and on $\overline{D}$ we define functions $x \mapsto x^\alpha$ and a space of polynomials of degree $\leq n$ by

$$\begin{array}{l} x^\alpha := \prod_{i=1}^{d} x_i^{\alpha_i} \\ P_n := \mathrm{span}\{x^\alpha : |\alpha \leq n\}. \end{array} \tag{2.5.47}$$

Setting $g = rg^*$, $r > 0$, we observe that

$$Ax^\alpha = \Big(c\sum_{i=1}^{d}(\theta_i - x_i)\tfrac{\partial}{\partial x_i} + \tfrac{r}{d}(1 - \sum_{j=1}^{d} x_j^2)\sum_{i=1}^{d}\tfrac{\partial^2}{\partial x_i{}^2}\Big)x^\alpha. \tag{2.5.48}$$

Since $\frac{\partial}{\partial x_i}x^\alpha \in P_{|\alpha|-1}$ and $\frac{\partial^2}{\partial x_i{}^2}x^\alpha \in P_{|\alpha|-2}$, we have

$$Ax^\alpha \in P_{|\alpha|} \qquad \forall \alpha \in A. \tag{2.5.49}$$

The spaces P_n are finite-dimensional and closed under A, and their union $\bigcup_n P_n$ is dense in $\mathcal{C}(\overline{D})$. Applying [16], Proposition 1.3.5, we see that A is closable and that its closure generates a Feller semigroup on $\mathcal{C}(\overline{D})$. This implies that the martingale problem is well-posed for A ([16], Theorem 4.4.1), and hence $(rg^*)_{r>0} \subset \mathcal{H}'$. But $F_c rg^* = (cr)/(c+r)g^*$, so the family $(rg^*)_{r>0}$ is closed under F_c for all $c \in (0, \infty)$. This implies that $rg^* \in \mathcal{H}''$ for all $r > 0$. ∎

Chapter 3

Clustering of Linearly Interacting Diffusions and Universality of their Long-Time Distribution

Abstract

We study infinite systems of diffusions indexed by an Abelian group Λ and taking values in a compact convex set $K \subset \mathbb{R}^d$ $(d \geq 1)$. Each diffusion is subject to: (1) a linear drift towards diffusions at surrounding positions, weighted with an interaction kernel $a(\cdot)$ on Λ, and (2) a diffusion with local rate $\sigma(\cdot)$ on K. For one-dimensional K, it is known that the system clusters (that is, becomes locally flat) if and only if the random walk on Λ with symmetrized kernel $a_S(\cdot) := a(\cdot)+a(-\,\cdot)$ is recurrent. We investigate the generalization of this statement to higher-dimensional K, focusing on a comparison argument that has been used in the one-dimensional case. We show that this argument is linked to the universality of the long-time distribution of the system, within the class of recurrent interaction kernels a_S, and this universality is in turn shown to follow from a condition involving the harmonic functions of the system. Under this condition we prove that the system clusters and we determine its long-time distribution. We give a general formula for certain special diffusion matrices that have previously appeared in the renormalization of the system, and we argue that universality properties found in this renormalization analysis find their origin in the same condition on the harmonic functions that we use.

3.1 Introduction and main results

3.1.1 Definitions

We consider models of linearly interacting diffusion processes. Models of this type were introduced in population biology and have been the subject of a considerable amount of mathematical work. We consider a family

$$X = (X_i)_{i\in\Lambda} = (X_i(t))_{t\geq 0,\, i\in\Lambda} \tag{3.1.1}$$

of stochastic processes, solving a system of stochastic differential equations of the following type:

$$dX_i(t) = \sum_{j\in\Lambda} a(j-i)(X_j(t) - X_i(t))dt + \sigma(X_i(t))dB_i(t) \qquad (i \in \Lambda,\ t \geq 0). \tag{3.1.2}$$

Here the following definitions apply.

- The $(B_i)_{i\in\Lambda}$ are standard d-dimensional Brownian motions, independent of each other and of the initial condition $X(0)$.

- The index set Λ is a finite or countable Abelian group, with

$$\begin{aligned} \text{group product} \quad & i+j \\ \text{inverse} \quad & -i \\ \text{unit element} \quad & 0. \end{aligned} \tag{3.1.3}$$

  For example, Λ may be the n-dimensional integer lattice $\mathbb{Z}^n$ or the N-dimensional hierarchical group Ω_N (as in [1], [10] and [21]). We sometimes refer to $i + j$ as addition and to 0 as the origin.

- The *interaction kernel* $a : \Lambda \to \mathbb{R}$ satisfies

$$\begin{aligned} & a(i) \geq 0 \qquad (i \in \Lambda) \\ & \sum_{i\in\Lambda} a(i) < \infty. \end{aligned} \tag{3.1.4}$$

  It is the kernel of a continuous-time random walk on Λ that jumps from a point i to a point j with rate $a(j-i)$. We assume that this random walk is irreducible.

- Each *single component* $X_i(t)$ takes values in a *state space* K that is a non-empty convex compact subset of $\mathbb{R}^d$. Thus, each component $X_i(t)$ itself consists of d components:

$$X_i(t) = (X_i^1(t), \ldots, X_i^d(t)). \tag{3.1.5}$$

Equation (3.1.2) componentwise reads

$$dX_i^\alpha(t) = \sum_j a(j-i)(X_j^\alpha(t) - X_i^\alpha(t))dt + \sum_\beta \sigma_{\alpha\beta}(X_i(t))dB_i^\beta(t)$$
$$(i \in \Lambda,\ \alpha = 1, \dots, d,\ t \geq 0). \tag{3.1.6}$$

We adopt the convention that sums over Roman indices $i, j, k, \dots$ range over Λ, while sums over Greek indices $\alpha, \beta, \gamma, \dots$ range from 1 to d.

- The function σ is a continuous function from K into $\mathbb{R}^d \otimes \mathbb{R}^d$, the space of $d \times d$ real matrices. It is a root of the *diffusion matrix* $w : K \to \mathbb{R}^d \otimes \mathbb{R}^d$:

$$w_{\alpha\beta}(x) := \tfrac{1}{2} \sum_\gamma \sigma_{\alpha\gamma}(x)\sigma_{\beta\gamma}(x). \tag{3.1.7}$$

We assume that w satisfies

$$\sum_{\alpha,\beta} z^\alpha w_{\alpha\beta}(x) z^\beta = 0 \quad \forall x \in K,\ \ z \in I_x^\perp, \tag{3.1.8}$$

where $I_x^\perp$ is the space of vectors perpendicular to

$$I_x := \{y \in \mathbb{R}^d : \exists \varepsilon > 0 \text{ such that } x + \lambda y \in K\ \forall |\lambda| \leq \varepsilon\}. \tag{3.1.9}$$

I_x is the space of directions in which the boundary of K at x is flat. In terms of the process X, condition (3.1.8) guarantees that the components $X_i(t)$ cannot leave the state space K.

We equip the space K^Λ with the product topology. In this topology K^Λ is a compact separable metrizable space. $\mathcal{C}(K^\Lambda)$ is the Banach space of continuous real-valued functions on K^Λ, equipped with the supremum norm $\|\cdot\|_\infty$.

Solutions to (3.1.2), whenever they exist, are continuous K^Λ-valued processes that solve the martingale problem for a linear operator A on $\mathcal{C}(K^\Lambda)$ given by

$$(Af)(x) := \Big(\sum_{ij} \sum_\alpha a(j-i)(x_j^\alpha - x_i^\alpha)\frac{\partial}{\partial x_i^\alpha} + \sum_i \sum_{\alpha\beta} w_{\alpha\beta}(x_i)\frac{\partial^2}{\partial x_i^\alpha \partial x_i^\beta} \Big) f(x). \tag{3.1.10}$$

Here f is a real function on K^Λ, and a typical element $x \in K^\Lambda$ is written as

$$x = (x_i)_{i\in\Lambda} = (x_i^\alpha)_{i\in\Lambda}^{\alpha=1,\dots,d}. \tag{3.1.11}$$

The operator A in (3.1.10) has domain

$$\mathcal{D}(A) := \mathcal{C}^2_{\text{fin}}(K^\Lambda), \tag{3.1.12}$$

the space of all $\mathcal{C}^2$-functions depending on finitely many coordinates only. For such functions the infinite sums of derivatives in (3.1.10) reduce to finite sums. Condition (3.1.8) guarantees that the operator A in (3.1.10) satisfies the maximum principle.

3.1.2 Existence and uniqueness: Theorems 3.1.1 and 3.1.2

We focus our attention on shift-invariant solutions to (3.1.2). For $j \in \Lambda$, let the shift operator $T_j : K^\Lambda \to K^\Lambda$ be defined as

$$(T_j x)_i := x_{i-j}. \tag{3.1.13}$$

We say that a solution X to (3.1.2) is shift-invariant if for each $j \in \Lambda$ the processes $(X(t))_{t\geq 0}$ and $(T_j X(t))_{t\geq 0}$ have the same finite-dimensional distributions. We say that a probability measure μ on K^Λ (equipped with the product-σ-field) is shift-invariant if $\mu = \mu \circ T_j^{-1}$ for all $j \in \Lambda$.

Theorem 3.1.1 *For each probability measure μ on K^Λ, there exists a solution $(X(t))_{t\geq 0}$ to (3.1.2) with initial condition $\mathcal{L}(X(0)) = \mu$ and sample paths in the continuous functions from $[0, \infty)$ to K^Λ. If μ is shift-invariant, then (3.1.2) has a shift-invariant solution with the same properties.*

If solutions to (3.1.2) are weakly unique, then any solution with a shift-invariant initial condition must be shift-invariant. Unfortunately, it is at present not very well understood when weak uniqueness holds for (3.1.2). Standard techniques give:

Theorem 3.1.2 *Assume that the function $\sigma : K \to \mathbb{R}^d \otimes \mathbb{R}^d$ is Lipschitz continuous. Then, for each K^Λ-valued initial condition $X(0)$, strong uniqueness holds for equation (3.1.2).*

Strong uniqueness for (3.1.2) implies weak uniqueness, just as in the case of finite-dimensional stochastic differential equations [39]. Theorem 3.1.2 does not cover many interesting cases. For example, for the Wright-Fisher diffusion matrix (see (3.1.15)), no root σ of w exists that satisfies the conditions of Theorem 3.1.2. For uniqueness results in this and a few more special cases, see [36, 37, 39]. In what follows, we avoid problems of uniqueness by assuming only the existence of solutions to (3.1.2).

3.1.3 Biological background

In population biology, models of the form (3.1.2) are used to describe the genetic composition of a population of individuals as a function of time. It is supposed that the population is divided into colonies, each containing a large number of individuals. A component $X_i(t) \in K$ describes the genetic composition of the population in colony i at time t. A typical choice for K is

$$K = S^{p-1} := \{(x^1, \ldots, x^{p-1}) : x^\alpha \geq 0 \;\forall \alpha, \; \sum_\alpha x^\alpha \leq 1\}. \tag{3.1.14}$$

For $p = 2$ we have a '2-type model'. In such a model a gene comes in two types ('alleles'), say type I and II. $X_i(t) \in S^1 = [0, 1]$ is interpreted as the relative frequency of type I in colony i at time t, the relative frequency of type II being $1 - X_i(t)$. More generally, in a 'p-type model' one considers the relative frequency of p types. $X_i^\alpha(t)$ is the relative frequency of the α-th type ($\alpha = 1, \dots, p-1$), the relative frequency of the remaining p-th type being $1 - \sum_\alpha X_i^\alpha(t)$.

The genetic compositions $X_i(t)$ change in time due to migration and resampling. Individuals in the population migrate between colonies according to a continuous-time random walk, jumping from j to i with rate $a(j-i)$. This migration causes an attractive interaction between components, expressed by the drift term $\sum_j a(j-i)(X_j(t) - X_i(t))dt$ in equation (3.1.2). At each colony after an exponential waiting time individuals are replaced by individuals of a type chosen at random from the colony. This resampling is expressed by the diffusion term $\sigma(X_i(t))dB_i(t)$ in equation (3.1.2), where σ is any continuous root of the diffusion matrix w. A typical choice for w is the

$$\text{Wright-Fisher diffusion matrix} \quad w_{\alpha\beta}(x) = x^\alpha(\delta_{\alpha\beta} - x^\beta). \tag{3.1.15}$$

For other choices, see the models listed in [5] and [38]. Often one's aim is to prove statements for as wide a class of diffusion matrices as possible. The following examples are found in the literature.

2-type models References [1, 10, 23] are concerned with diffusion functions $w : [0, 1] \to [0, \infty)$, Lipschitz continuous, satisfying

$$w(x) = 0 \Leftrightarrow x \in \{0, 1\}. \tag{3.1.16}$$

Isotropic models Reference [21] is concerned with diffusion matrices of the form

$$w_{\alpha\beta}(x) = \delta_{\alpha\beta} g(x), \tag{3.1.17}$$

where $g : K \to [0, \infty)$ is a nice function satisfying

$$g(x) = 0 \Leftrightarrow x \in \partial K, \tag{3.1.18}$$

with ∂K the (topological) boundary of K.

Work on the non-compact state space $K = [0, \infty)$ can be found in [2, 11]. Reference [14] is concerned with an isotropic model on $K = [0, \infty)^2$. A generalization of the p-type model to infinitely many types is studied in [12].

3.1.4 The non-interacting model

Later on, we will make a comparison between the model in (3.1.2) and a model without interaction. For this, we consider the case that the Abelian group Λ consists of only one element. Equation (3.1.2) now reduces to

$$dX(t) = \sigma(X(t))dB(t) \qquad (t \geq 0), \tag{3.1.19}$$

where $(X(t))_{t\geq 0}$ is a K-valued stochastic process. Uniqueness of solutions to (3.1.19) can be proved under considerably weaker conditions than those needed for equation (3.1.2) (see the examples in section 3.1.10). We therefore prefer, when possible, to assume only existence of solutions to (3.1.2) and uniqueness of solutions to (3.1.19).

Solutions to (3.1.19) are bounded martingales, and hence they have a last element:

$$X^x(t) \to X^x(\infty) \ \text{ a.s.} \quad \text{as } t \to \infty \quad (x \in K), \tag{3.1.20}$$

where X^x is the solution of (3.1.19) starting in x. $X^x(\infty)$ takes values in the set

$$\partial_w K := \{x \in K : w_{\alpha\beta}(x) = 0 \ \forall \alpha, \beta\}. \tag{3.1.21}$$

In typical examples, $\partial_w K$ is a subset of the (topological) boundary of K. We call $\partial_w K$ the *effective boundary* of K. We denote the law of $X^x(\infty)$ by

$$\Gamma_x := \mathcal{L}(X^x(\infty)). \tag{3.1.22}$$

The collection $(\Gamma_x)_{x\in K}$ we call the *boundary distribution* associated with the diffusion matrix w. We note that different diffusion matrices w may share the same boundary distribution. For example, by the martingale property of solutions to (3.1.19), diffusions on $[0, 1]$ with w as in (3.1.16) all have

$$\Gamma_x = (1 - x)\delta_0 + x\delta_1. \tag{3.1.23}$$

For diffusions with isotropic w as in (3.1.17), solutions to (3.1.19) are time-transformed Brownian motions, and therefore

$$\Gamma_x = \mathcal{L}(B^x_\tau), \tag{3.1.24}$$

where $(B^x_t)_{t\geq 0}$ is Brownian motion starting in x and

$$\tau := \inf\{t \geq 0 : B_t \in \partial K\}. \tag{3.1.25}$$

We try to answer two questions. When does the distribution of components $X_i(t)$ of the interacting system in (3.1.2) converge to a distribution on the effective boundary $\partial_w K$? And when is this limiting distribution actually the same as in the non-interacting system?

3.1.5 Clustering: Theorem 3.1.3

In order to state our first result, we introduce the *symmetrized kernel*

$$a_S(i) := a(i) + a(-i) \qquad (i \in \Lambda). \tag{3.1.26}$$

By the random walk with kernel a_S we mean a continuous-time random walk on Λ that jumps from a point j to a point i with rate $a_S(j-i)$. By $\Rightarrow$ we denote weak convergence of the laws of processes as probability measures on K^Λ.

Theorem 3.1.3 *Let X be a shift-invariant solution to (3.1.2) and assume that there exists a K^Λ-valued random variable $X(\infty)$ such that*

$$X(t) \Rightarrow X(\infty) \qquad \textit{as } t \to \infty. \tag{3.1.27}$$

If the random walk with kernel a_S is recurrent, then

$$\begin{array}{ll} (i) & P[X_i(\infty) \in \partial_w K \ \ \forall i \in \Lambda] = 1 \\ (ii) & P[X_i(\infty) = X_j(\infty) \ \ \forall i, j \in \Lambda] = 1. \end{array} \tag{3.1.28}$$

If the random walk with kernel a_S is transient, $E[X_0(0)] \notin \partial_w K$ and $\mathcal{L}(X(0))$ is spatially ergodic, then

$$\begin{array}{lll} (i) & P[X_i(\infty) \in \partial_w K] < 1 & \forall i \in \Lambda \\ (ii) & P[X_i(\infty) = X_j(\infty)] < 1 & \forall i \neq j \in \Lambda. \end{array} \tag{3.1.29}$$

Note that Theorem 3.1.3 makes a statement about the possible properties of a limiting distribution $X(\infty)$, but that it does not answer the question whether such a limiting distribution actually exists. Provided we know in some way that $X(t)$ converges weakly to a limit, Theorem 3.1.3 says the following.

In the *recurrent* case, the configuration in any finite window $\Delta \subset \Lambda$ after a sufficiently long time becomes almost flat. At large but finite time in the system there are regions, called 'clusters', of typical sizes that grow with time, in which all components are almost equal. This behavior is called 'clustering'. The behavior is similar to that of the voter model in low ($d \leq 2$) dimension. In fact, 2-type models as in (3.1.16) are believed to be asymptotically equivalent, in some sense, to the voter model on the same lattice. See [7] for some pictures of simulations of the (clustering) voter model on $\mathbb{Z}^2$.

In the *transient* case, such clustering behavior cannot occur. Instead, the system converges to a 'true' equilibrium $X(\infty)$. We refer to this as 'stable' behavior.

Although it seems hard to imagine a shift-invariant solution to (3.1.2) that does not converge as $t \to \infty$, the convergence in (3.1.27) is in general hard to prove. For finite Λ, one may exploit the fact that $\sum_i X_i(t)$ is a bounded martingale to get the convergence in (3.1.27), not only in the sense of weak convergence, but also in L^2-norm.[1] For infinite Λ, convergence in L^2-norm in general does not hold.

[1] For the interested reader we have added a proof of this fact in section 3.6.

For the 2-type model, the convergence in (3.1.27) has been proved in [6, 26]. In [6], this is achieved for transient a_S by a coupling technique, and for recurrent a_S by a 'duality comparison argument'. This argument, as well as Theorem 3.1.3, are based on calculations involving covariances between components.

3.1.6 Covariance calculations: Lemma 3.1.4

In this section we explain the relation between covariance calculations, the random walk with kernel a_S, and clustering properties of the system in (3.1.2).

For any two K-valued random variables X and Y the covariance of X and Y is the quantity

$$\mathrm{Cov}(X, Y) = E[X \cdot Y] - E[X] \cdot E[Y], \tag{3.1.30}$$

where $\cdot$ denotes the inner product $x \cdot y = \sum_\alpha x^\alpha y^\alpha$. By $tr(w)$ we denote the trace

$$tr(w)(x) = \sum_{\alpha=1}^{d} w_{\alpha\alpha}(x) \qquad (x \in K) \tag{3.1.31}$$

of the diffusion matrix w. The following lemma follows from a little calculation involving Itô's formula and a bit of continuity.

Lemma 3.1.4 *Let X be a shift-invariant solution of (3.1.2). Then there exists a $\theta \in K$ such that*

$$E[X_i(t)] = \theta \qquad (t \geq 0,\ i \in \Lambda) \tag{3.1.32}$$

and there exists a function $C : [0, \infty) \times \Lambda \to \mathbb{R}$ such that

$$\mathrm{Cov}(X_i(t), X_j(t)) = C_t(j - i) \qquad (t \geq 0,\ i, j \in \Lambda). \tag{3.1.33}$$

For each i, the function $t \mapsto C_t(i)$ is continuously differentiable and satisfies

$$\tfrac{\partial}{\partial t} C_t(i) = \sum_j a_S(j - i)(C_t(j) - C_t(i)) + 2\delta_{i0} E[tr(w)(X_0(t))]. \tag{3.1.34}$$

The right-hand side of (3.1.34) contains the operator

$$(Gf)(i) := \sum_j a_S(j - i)(f(j) - f(i)), \tag{3.1.35}$$

acting on bounded functions $f : \Lambda \to \mathbb{R}$. G is the generator of the random walk with kernel a_S. For solutions to (3.1.34) we have the representation

$$C_t(i) = \sum_j P_t(j - i) C_0(j) + 2 \int_0^t P_s(0 - i) E[tr(w)(X_0(t - s))] ds, \tag{3.1.36}$$

where $P_t(j-i)$ is the probability that the random walk with kernel a_S starting from a point i, is in j at time t.

In view of the biological background of the model, the representation in (3.1.36) can be understood in terms of a 'historical process' tracing back where ancestors of two individuals from colonies at 0 and i lived at previous times. The time the symmetrized random walk spends at the origin is the time the ancestors lived in the same colony, and hence had a chance of descending from a common ancestor.

This sort of reasoning works best when w is the Wright-Fisher diffusion matrix. In that case the system (3.1.2) is in duality with a system of delayed coalescing random walks (see formula (4.1) in [23] or Lemma 2.3 in [34]) and all mixed moments of the type $E[X_i(t)X_j(t)]$, $E[X_i(t)X_j(t)X_k(t)]$,... may be expressed in terms of the dual model. This duality has been exploited in [34] to show the dichotomy between clustering and stable behavior for the Wright-Fisher diffusion on [0, 1].

For arbitrary w, the representation (3.1.36) is sufficient to derive Theorem 3.1.3, but not to derive the convergence in (3.1.27). For 2-type models as in (3.1.16), this shortcoming can be overcome by using a 'duality comparison argument' as in [6] (see also [5]), which makes a comparison between models with arbitrary w and the special model with Wright-Fisher diffusion, for which clustering can be derived using duality.

3.1.7 Universality of the long-time distribution: Theorem 3.1.5

We give sufficient conditions for the convergence in (3.1.27) and for the uniqueness in distribution of the limit $X(\infty)$. For this we need to look at the differential equation

$$dY(t) = (\theta - Y(t))dt \qquad (t \geq 0), \tag{3.1.37}$$

where $\theta \in K$ is a fixed parameter. By the convexity of K, the solution of (3.1.37) starting from a point $x \in K$:

$$Y^x(t) = \theta + (x-\theta)e^{-t} \qquad (t \geq 0), \tag{3.1.38}$$

stays in K for all time. Solutions to (3.1.37) are associated with a semigroup $(T_{\theta,t})_{t\geq 0}$ on the space $B(K)$ of bounded measurable real functions on K, given by

$$(T_{\theta,t}f)(x) := E[f(Y^x(t))] = f(\theta+(x-\theta)e^{-t}) \qquad (x, \theta \in K,\ t \geq 0). \tag{3.1.39}$$

We are going to compare equation (3.1.37) (non-zero drift, zero diffusion) with the non-interacting equation (3.1.19) (zero drift, non-zero diffusion).

Let us assume that for each initial condition $x \in K$, the non-interacting equation (3.1.19) has a unique weak solution $(X^x(t))_{t\geq 0}$, and let us denote the associated semigroup on $B(K)$ by

$$(S_t f)(x) := E[f(X^x(t))] \qquad (x \in K,\ t \geq 0). \tag{3.1.40}$$

We add a 'last element' S_∞ to this semigroup by defining

$$(S_\infty f)(x) := E[f(X^x(\infty))] = \int_K \Gamma_x(dy) f(y) \qquad (x \in K,\ f \in B(K)), \tag{3.1.41}$$

where $(\Gamma_x)_{x\in K}$ is the boundary distribution associated with w, introduced in (3.1.22).

With this notation, we formulate a condition that will guarantee that the long-time behavior of the non-interacting model is not changed by the introduction of a linear drift.

Definition 3.1.1 *Let w be a diffusion matrix on K such that weak uniqueness holds for (3.1.19), and let $(\Gamma_x)_{x\in K}$ be the associated boundary distribution. We say that $(\Gamma_x)_{x\in K}$ is stable against a linear drift if*

$$S_\infty T_{\theta,t} S_\infty f = T_{\theta,t} S_\infty f \qquad \forall \theta \in K,\ t \geq 0,\ f \in B(K). \tag{3.1.42}$$

Since $S_\infty S_\infty = S_\infty$, we can read equation (3.1.42) as: S_∞ and $T_{\theta,t}$ commute on functions of the form $S_\infty f$.

For technical reasons, we will restrict ourselves to the case that

$$S_\infty(\mathcal{C}(K)) \subset \mathcal{C}(K). \tag{3.1.43}$$

This condition guarantees that $S_\infty f$ is a w-harmonic function for all $f \in \mathcal{C}(K)$, where the space of w-harmonic functions is defined as

$$H := \{f \in \mathcal{D}(G) : Gf = 0\}, \tag{3.1.44}$$

with G the full generator of the process in (3.1.19) and $\mathcal{D}(G)$ its domain. In particular, $\mathcal{C}^2$-functions are w-harmonic if and only if they solve the equation

$$\sum_{\alpha\beta} w_{\alpha\beta}(x) \tfrac{\partial^2}{\partial x^\alpha \partial x^\beta} f(x) = 0 \qquad (x \in K). \tag{3.1.45}$$

It turns out that condition (3.1.42) is equivalent to

$$T_{\theta,t}(H) \subset H \qquad \forall \theta \in K,\ t \geq 0. \tag{3.1.46}$$

That is, for each θ the space of w-harmonic functions is invariant under the semigroup $(T_{\theta,t})_{t\geq 0}$.

With these definitions, our main result reads as follows.

Theorem 3.1.5 *Let X be a shift-invariant solution to (3.1.2) such that $\mathcal{L}(X(0))$ is spatially ergodic and*

$$E[X_i(0)] = \theta \qquad (i \in \Lambda) \tag{3.1.47}$$

for some $\theta \in K$. Assume that weak uniqueness holds for the non-interacting equation (3.1.19), that the associated boundary distribution is stable against a linear drift, that $S_\infty(\mathcal{C}(K)) \subset \mathcal{C}(K)$ and that H is contained in the bp-closure of $\mathcal{C}^2(K) \cap H$. If the random walk with kernel a_S is recurrent, then there exists a K^Λ-valued random variable $X(\infty)$ such that

$$X(t) \Rightarrow X(\infty) \qquad \text{as } t \to \infty, \tag{3.1.48}$$

where

$$\mathcal{L}(X_i(\infty)) = \Gamma_\theta \qquad (i \in \Lambda). \tag{3.1.49}$$

The bp-closure of a set is the smallest set containing it that is closed under bounded pointwise limits.

Note that by Theorem 3.1.3, $P[X_i(\infty) = X_j(\infty)\ \forall i, j \in \Lambda] = 1$. Thus, the fact that the boundary distribution is stable against a linear drift not only allows us to conclude that $X(t)$ converges to a limit $X(\infty)$, it also allows us to completely specify its distribution. This distribution turns out to be *universal* in all recurrent random walk kernels a_S and Abelian groups Λ, and in all diffusion matrices w sharing the same boundary distribution $(\Gamma_x)_{x\in K}$.

3.1.8 Harmonic functions: Lemma 3.1.6

To see what goes into proving Theorem 3.1.5, we mention the following:

Lemma 3.1.6 *Let X be a solution to (3.1.2). Assume that weak uniqueness holds for the non-interacting equation (3.1.19), that the associated boundary distribution is stable against a linear drift, that $S_\infty(\mathcal{C}(K)) \subset \mathcal{C}(K)$ and that H is contained in the bp-closure of $\mathcal{C}^2(K) \cap H$. Then*

$$E[f(X_i(t))] = E\Big[f\Big(\sum_j P_t(j-i)X_j(0)\Big)\Big] \qquad \forall f \in H,\ i \in \Lambda,\ t \geq 0, \tag{3.1.50}$$

where $P_t(j-i)$ is the probability that the random walk with kernel a starting from i, is in j at time t.

The situation is particularly simple when $X_i(0) = \theta$ for all $i \in \Lambda$. In that case

$$E[f(X_i(t))] = f(\theta) \qquad \forall f \in H,\ i \in \Lambda,\ t \geq 0. \tag{3.1.51}$$

For a 2-type model with diffusion matrix w as in (3.1.16), the class H contains only affine functions $x \mapsto a + bx$ ($a, b \in \mathbb{R}$), and (3.1.51) says no more than that the mean of the components is conserved. Since there is only one distribution on $\{0, 1\}$ with a given mean, it is then immediately clear (for recurrent a_S) that there is only one possible long-time distribution for the process in (3.1.2). In the general higher-dimensional case, we need to specify a distribution on the effective boundary $\partial_w K$, and for this we need the expectation of sufficiently many harmonic functions. We may describe (3.1.51) by saying that the 'w-harmonic mean' of the components is conserved.

3.1.9 Special models: Corollary 3.1.7

The proof of Theorem 3.1.5 uses a comparison argument, in the spirit of the 'duality comparison argument' in [6]. In our comparison argument we use objects related to the *special diffusion matrix*

$$w^*_{\alpha\beta}(x) := \int_K \Gamma_x(dy)(y^\alpha - x^\alpha)(y^\beta - x^\beta) \qquad (x \in K, \alpha, \beta = 1, \dots, d). \tag{3.1.52}$$

We do not have a duality for the model with w^*, but we can find an expression for second moments, which is enough for our purposes. For the special model with $w = \lambda w^*$ ($\lambda > 0$) the proof of Theorem 3.1.5 yields the following corollary.

Corollary 3.1.7 *In addition to the assumptions in Theorem 3.1.5, assume that for some $\lambda \in (0, \infty)$*

$$w(x) = \lambda w^*(x) \qquad (x \in K). \tag{3.1.53}$$

Then for each $t \geq 0$, $i, j \in \Lambda$, $\alpha, \beta = 1, \dots, d$

$$E[(X^\alpha_i(t) - \theta^\alpha)(X^\beta_j(t) - \theta^\beta)] = w^*_{\alpha\beta}(\theta) K^\lambda_t(i, j), \tag{3.1.54}$$

where $K^\lambda_t(i, j)$ denotes the probability that two delayed coalescing random walks, each with kernel a, starting in points i and j respectively, and coalescing with rate 2λ, have coalesced before time t.

3.1.10 Examples

We close this introduction by giving two examples of classes of diffusion matrices w satisfying the assumptions in Theorem 3.1.5.

The first example arises when we generalize the 2-type models mentioned in (3.1.16) to p-type models in the following way.

Example 3.1.8 (p-type models) *Assume that K is the $(p-1)$-dimensional simplex S^{p-1}, and that $x \mapsto w(x)$ is Lipschitz continuous and satisfies (compare (3.1.8))*

$$\sum_{\alpha,\beta} z^\alpha w_{\alpha\beta}(x) z^\beta = 0 \quad \Leftrightarrow \quad z \in I_x^\perp \qquad (x \in K). \tag{3.1.55}$$

Then:
(a) *Weak uniqueness holds for the non-interacting equation (3.1.19). The boundary distribution is stable against a linear drift, $S_\infty(\mathcal{C}(K)) \subset \mathcal{C}(K)$, and H is contained in the bp-closure of $\mathcal{C}^2(H) \cap H$.*
(b) *The class of w-harmonic functions consists of all affine functions*

$$x \mapsto a + \sum_\alpha b^\alpha x^\alpha \qquad (a, b_1, \ldots, b_d \in \mathbb{R}). \tag{3.1.56}$$

(c) *The associated special diffusion matrix is the Wright-Fisher diffusion matrix*

$$w^*_{\alpha\beta}(x) = x^\alpha(\delta_{\alpha\beta} - x^\beta) \qquad (x \in K,\ \alpha, \beta = 1, \ldots, d). \tag{3.1.57}$$

The second example is formed by the class of isotropic diffusion matrices (compare (3.1.17)).

Example 3.1.9 (isotropic models) *Assume that K has non-empty interior K°, and let $\partial K := K \backslash K^\circ$ denote its topological boundary. Assume that*

$$w_{\alpha\beta}(x) = \delta_{\alpha\beta} g(x) \qquad (x \in K,\ \alpha, \beta = 1, \ldots, d) \tag{3.1.58}$$

for some Lipschitz continuous function $g : K \to [0, \infty)$ satisfying

$$g(x) = 0 \quad \Leftrightarrow \quad x \in \partial K. \tag{3.1.59}$$

Then:
(a) *Weak uniqueness holds for the non-interacting equation (3.1.19). The boundary distribution is stable against a linear drift, $S_\infty(\mathcal{C}(K)) \subset \mathcal{C}(K)$, and H is contained in the bp-closure of $\mathcal{C}^2(H) \cap H$.*
(b) *The class of w-harmonic functions is given by*

$$H = \Big\{ f \in \mathcal{C}(K) \cap \mathcal{C}^2(K^\circ) : \sum_\alpha \tfrac{\partial^2}{\partial x^{\alpha 2}} f(x) = 0 \text{ on } K^\circ \Big\}. \tag{3.1.60}$$

(c) *The associated special diffusion matrix is given by*

$$w^*_{\alpha\beta}(x) = \delta_{\alpha\beta} g^*(x) \qquad (x \in K,\ \alpha, \beta = 1, \ldots, d), \tag{3.1.61}$$

where $g^* \in \mathcal{C}(K) \cap \mathcal{C}^2(K^\circ)$ is the unique solution of

$$\begin{aligned} -\tfrac{1}{2} \sum_\alpha \tfrac{\partial^2}{\partial x^{\alpha 2}} g^*(x) &= 1 \qquad (x \in K^\circ) \\ g^*(x) &= 0 \qquad (x \in \partial K). \end{aligned} \tag{3.1.62}$$

One can find a few more examples of diffusion matrices satisfying the assumptions in Theorem 3.1.5, but it turns out that these are mainly trivial variations on the two examples mentioned above. The message of Theorem 3.1.5 is that all these examples fall into the same framework. The common property that unites them is the stability of the boundary distribution against a linear drift.

In fact, we conjecture that this property is a *necessary* condition for the universality of the long-time distribution. If the boundary distribution is not stable against a linear drift, it seems likely that still $X(t)$ converges weakly to some limit as $t \to \infty$, although we do not know how to prove this for infinite Λ. But we believe that in this case the law of $X(\infty)$ will depend on the choice of the recurrent kernel a_S and the Abelian group Λ. However, we have at present very little knowledge about the nature of this dependence.

In conclusion, we have found that the 'duality comparison argument' developed in [6] is linked to universality of the long-time distribution of solutions to (3.1.2). A similar relation between comparison arguments and universality has been found for models on the hierarchical group Ω_N with N large in [1, 10, 21]. There, the system in (3.1.2) is studied by means of a renormalization transformation acting on diffusion matrices. Under iteration of the transformation, the renormalized diffusion matrices converge to a limit. In the clustering case a comparison argument shows that this limit is universal within a large 'universality class' of matrices. This has been worked out for 2-type models in [1] and for isotropic models in [21]. The universal limit that is found is exactly the w^* in formula (3.1.52). The conclusion we can draw from Theorem 3.1.5 is that the correct 'universality classes' of diffusion matrices one should look at are formed by all diffusion matrices w that share the same boundary distribution $(\Gamma_x)_{x\in K}$. Furthermore, universal behavior can be expected only if this boundary distribution is stable against a linear drift.

3.2 Proofs of Theorems 3.1.1 and 3.1.2

3.2.1 Proof of Theorem 3.1.1

If $\Delta \subset \Lambda$ is finite then $\mathcal{C}^2(K^\Delta)$ is the space of real functions on K^Δ that have a $\mathcal{C}^2$-extension to all of $(\mathbb{R}^d)^\Delta$. $\mathcal{C}^2_{\text{fin}}(K^\Lambda)$ consists of all functions that are the lifting to the larger space K^Λ of a function in $\mathcal{C}^2(K^\Delta)$ for some finite $\Delta \subset \Lambda$.

Lemma 3.2.1 *The operator A in (3.1.10) with domain $\mathcal{D}(A)$ in (3.1.12) is a densely defined linear operator on the Banach space $\mathcal{C}(K^\Lambda)$, and satisfies the maximum principle.*

Proof of Lemma 3.2.1: By the Stone-Weierstrass theorem, $\mathcal{C}^2(K^\Delta)$ is dense in $\mathcal{C}(K^\Delta)$ for each finite $\Delta \subset \Lambda$. Pick a bijection between Λ and the positive integers and fix a point $z \in K$. Define, for $x \in K^\Lambda$,

$$\pi_n(x) := (x_1, x_2, \ldots, x_n, z, z, \ldots). \tag{3.2.1}$$

The sets $\pi(K^\Lambda)$ are uniformly dense in K^Λ, and since each $f \in \mathcal{C}(K^\Lambda)$ is uniformly continuous, it is the uniform limit of functions

$$f_n(x) := f(\pi_n(x)) \tag{3.2.2}$$

depending on finitely many coordinates. Hence $\mathcal{C}_{\text{fin}}(K^\Lambda)$ is dense in $\mathcal{C}(K^\Lambda)$.

To see that A satisfies the maximum principle, fix $f \in \mathcal{C}^2_{\text{fin}}(K^\Lambda)$ and suppose that f assumes its maximum in a point x. Fix an $i \in \Lambda$. Keeping all $(x_j)_{j \neq i}$ fixed, f assumes its maximum as a function of the remaining variable in the point x_i. By the convexity of K it is easily checked that

$$\sum_j \sum_\alpha a(j-i)(x_j^\alpha - x_i^\alpha)\tfrac{\partial}{\partial x_i^\alpha} f(x) \leq 0. \tag{3.2.3}$$

Condition (3.1.8) ensures that

$$\sum_{\alpha\beta} w_{\alpha\beta}(x_i)\frac{\partial^2}{\partial x_i^\alpha \partial x_i^\beta} f(x) \leq 0, \tag{3.2.4}$$

as can be seen by writing the matrix $w(x)$ in diagonal form:

$$\sum_{\alpha\beta} w_{\alpha\beta}(x_i)\frac{\partial^2}{\partial x_i^\alpha \partial x_i^\beta} f(x) = \sum_{\tilde\alpha} \lambda_{\tilde\alpha}(x)\frac{\partial^2}{\partial x^{\tilde\alpha 2}} f(x) \tag{3.2.5}$$

for an appropriate orthonormal basis $(x^{\tilde\alpha})$ of $\mathbb{R}^d$. By condition (3.1.8), the only non-zero terms in (3.2.5) occur for directions that lie in the space I_x, and for such directions the second derivative is non-positive. ∎

We equip K^Λ with the Borel σ-field generated by the open sets. We write $\mathcal{D}_{K^\Lambda}[0,\infty)$ for the cadlag functions from $[0,\infty)$ to K^Λ, equipped with the metric d from chapter 3, section 5 of [16], which generates the Skorohod topology. By $\mathcal{C}_{K^\Lambda}[0,\infty)$ we denote the continuous functions from $[0,\infty)$ to K^Λ. On $\mathcal{D}_{K^\Lambda}[0,\infty)$ we choose the Borel σ-field generated by the open sets of this topology. We equip the space of probability measures on $\mathcal{D}_{K^\Lambda}[0,\infty)$ with the topology of weak convergence and we denote weak convergence of the laws of processes with sample paths in $\mathcal{D}_{K^\Lambda}[0,\infty)$ by $\Rightarrow$. Thus $X_n \Rightarrow X$ means that

$$E[f(X_n)] \to E[f(X)] \quad \text{as } n \to \infty \tag{3.2.6}$$

for all bounded continuous real functions f on $\mathcal{D}_{K^\Lambda}[0,\infty)$. By a solution to the martingale problem we always mean a solution with sample paths in $\mathcal{D}_{K^\Lambda}[0,\infty)$.

Lemma 3.2.2 *For each probability measure on K^Λ there exists a solution to the martingale problem for A with initial condition μ. Each solution to the martingale problem for A has sample paths in $\mathcal{C}_{K^\Lambda}[0,\infty)$. The space of solutions to the martingale problem for A is compact in the topology of weak convergence. If X_n, X solve the martingale problem for A, then $X_n \Rightarrow X$ implies $X_n(t) \Rightarrow X(t)$ for all $t \geq 0$.*

Proof of Lemma 3.2.2: Existence of solutions to the martingale problem for A follows from Lemma 3.2.1 in combination with Theorem 5.4 and Remark 5.5 from chapter 4 of [16].

The continuity of sample paths can be shown by Problem 19 from the same chapter: for this one needs to find for every $x \in K^\Lambda$ a function $f_x \in \mathcal{D}(A)$ such that for every $\varepsilon > 0$

$$\inf\{f_x(y) - f_x(x) : x, y \in K^\Lambda,\ d(x,y) \geq \varepsilon\} > 0 \tag{3.2.7}$$

and such that $\lim_{x\to y} Af_x(y) = Af_y(y) = 0$ for all $x \in K^\Lambda$. Instead of working with A, one may also use the closure of A. Applying Lemma 3.4.5 below and defining $(\gamma_i)_{i\in\Lambda}$ as in (3.2.14), it is not hard to check that the functions

$$f_x(y) := \sum_i \gamma_i |x_i - y_i|^3 \tag{3.2.8}$$

satisfy the requirements.

Compactness of the space of solutions follows from Lemma 5.1 and Remark 5.2 from chapter 4 of [16]. Finally, weak convergence in path space of solutions X_n to the martingale problem for A implies convergence of finite-dimensional distributions by Theorem 7.8 from chapter 3 of [16] and the continuity of sample paths. ∎

Proof of Theorem 3.1.1: Existence of solutions to the martingale problem for A is guaranteed by Lemma 3.2.2. Corollary 3.4 from chapter 5 of [16] generalizes in a straightforward way to the infinite-dimensional case, and so for each solution to the martingale problem for A we can find a weak solution to the stochastic differential equation (3.1.2).

We next show that for each shift-invariant initial condition μ, there exists a shift-invariant solution to (3.1.2). It suffices to construct a shift-invariant solution to the martingale problem for A. We define a shift operation on $\mathcal{D}_{K^\Lambda}[0,\infty)$ in the obvious way, by putting

$$(\mathcal{T}_j x)_i(t) := x_{i-j}(t) \qquad (i, j \in \Lambda,\ t \geq 0). \tag{3.2.9}$$

Let X be a solution to the martingale problem for A with initial condition $\mathcal{L}(X(0)) = \mu$. By Lemma 3.3.3 below, there exists a sequence of functions $p_n : \Lambda \to [0, \infty)$ such that $\sum_i p_n(i) = 1$ for each n and

$$\lim_{n\to\infty} \sum_k |p_n(i-k) - p_n(j-k)| = 0 \qquad \forall i, j \in \Lambda. \tag{3.2.10}$$

Let (X_n) be a sequence of processes with sample paths in $\mathcal{D}_{K^\Lambda}[0, \infty)$ with law

$$\mathcal{L}(X_n) = \sum_k p_n(k)\mathcal{L}(\mathcal{T}_k X). \tag{3.2.11}$$

Then each X_n solves the martingale problem for A with initial condition $\sum_k p_n(k)(\mu \circ T_k^{-1}) = \mu$, where we use that μ is shift-invariant. By Lemma 3.2.2 we can find a subsequence $(X_{n(m)})$ and a solution X^∞ to the martingale problem for A such that $X_{n(m)} \Rightarrow X^\infty$. Clearly X^∞ has initial condition $\mathcal{L}(X^\infty(0)) = \mu$ and for any bounded continuous real function f on $\mathcal{D}_{K^\Lambda}[0, \infty)$ we have

$$\begin{aligned}
&|E[f(\mathcal{T}_j X_{n(m)})] - E[f(X_{n(m)})]| \\
&= \Big|\sum_k p_{n(m)}(k)E[f(\mathcal{T}_j\mathcal{T}_k X)] - \sum_i p_{n(m)}(k)E[f(\mathcal{T}_k X)]\Big| \\
&= \Big|\sum_k p_{n(m)}(k-j)E[f(\mathcal{T}_k X)] - \sum_i p_{n(m)}(k)E[f(\mathcal{T}_k X)]\Big| \\
&\leq \sum_k |p_{n(m)}(k-j) - p_{n(m)}(k)|\, \|f\|_\infty.
\end{aligned} \tag{3.2.12}$$

By (3.2.10) it follows that $\mathcal{T}_j X^\infty$ and X^∞ have the same distribution as a probability measure on $\mathcal{D}_{K^\Lambda}[0, \infty)$, which implies that their finite-dimensional distributions agree. Hence X^∞ is shift-invariant. ∎

3.2.2 Proof of Theorem 3.1.2

Define a normalized interaction kernel $\tilde{a}$ and a normalizing constant Z by

$$Z := \sum_i a(i) \qquad \tilde{a}(i) := Z^{-1}a(i). \tag{3.2.13}$$

For each $M > 1$ there exist [38] strictly positive numbers $(\gamma_i)_{i\in\Lambda}$ such that $\sum_i \gamma_i < \infty$ and

$$\sum_i \tilde{a}(j-i)\gamma_i \leq M\gamma_j \quad (j \in \Lambda). \tag{3.2.14}$$

Let $L^2(\gamma)$ be the Hilbert space

$$L^2(\gamma) := \{x \in (\mathbb{R}^d)^\Lambda : \sum_i \gamma_i |x_i|^2 < \infty\} \tag{3.2.15}$$

with inner product

$$\langle x, y \rangle_\gamma := \sum_i \gamma_i \, x_i \cdot y_i, \tag{3.2.16}$$

where $\cdot$ denotes the standard inner product on $\mathbb{R}^d$. Clearly, $K^\Lambda \subset L^2(\gamma)$ and the topology on K^Λ coincides with the topology on $L^2(\gamma)$. We write $\|x\|_\gamma := \sqrt{\langle x, x\rangle_\gamma}$ for the Hilbert norm on $L^2(\gamma)$.

Set $\Delta(t) := X(t) - \tilde{X}(t)$, where X and $\tilde{X}$ are solutions to (3.1.2), starting in $X(0) = \tilde{X}(0)$ and adapted to the same set of Brownian motions. Then

$$\begin{aligned} d\Delta_i^\alpha(t) &= Z \sum_j \tilde{a}(j-i)(\Delta_j^\alpha(t) - \Delta_i^\alpha(t))dt \\ &\quad + \sum_\beta (\sigma_{\alpha\beta}(X_i(t)) - \sigma_{\alpha\beta}(\tilde{X}_i(t)))dB_i^\beta(t). \end{aligned} \tag{3.2.17}$$

By Itô's formula we see that

$$\begin{aligned} E\|\Delta(T)\|_\gamma^2 = \int_0^T E\Big\{ & 2\sum_i \sum_\alpha \gamma_i \Delta_i^\alpha(t) Z \sum_j \tilde{a}(j-i)(\Delta_j^\alpha(t) - \Delta_i^\alpha(t)) \\ & + \sum_i \gamma_i \sum_{\alpha\beta} (\sigma_{\alpha\beta}(X_i(t)) - \sigma_{\alpha\beta}(\tilde{X}_i(t)))^2 \Big\} dt. \end{aligned} \tag{3.2.18}$$

By the Lipschitz property of σ we have

$$\Big(\sum_{\alpha\beta} (\sigma_{\alpha\beta}(x) - \sigma_{\alpha\beta}(y))^2\Big)^{\frac{1}{2}} \leq L|x-y| \quad (x, y \in K) \tag{3.2.19}$$

for some $L < \infty$. With $(\tilde{a}\Delta(t))_i^\alpha := \sum_j \tilde{a}(j-i)\Delta_j^\alpha(t)$ it follows that

$$\begin{aligned} E\|\Delta(T)\|_\gamma^2 &\leq \int_0^T E\Big\{2Z\langle \Delta(t), \tilde{a}\Delta(t) - \Delta(t)\rangle_\gamma + L^2\|\Delta(t)\|_\gamma^2\Big\}dt \\ &\leq \int_0^T E\Big\{2Z(\|\Delta(t)\|_\gamma \|\tilde{a}\Delta(t)\|_\gamma - \|\Delta(t)\|_\gamma^2) + L^2\|\Delta(t)\|_\gamma^2\Big\}dt \\ &\leq \int_0^T (2Z(M^{\frac{1}{2}} - 1) + L^2) E\|\Delta(t)\|_\gamma^2 dt, \end{aligned} \tag{3.2.20}$$

where we used Cauchy-Schwarz and the fact that, by Jensen's inequality and by (3.2.14),

$$
\begin{aligned}
\|\tilde{a}x\|_\gamma^2 &= \sum_i \gamma_i \Big| \sum_j \tilde{a}(j-i)x_j \Big|^2 \leq \sum_{ij} \gamma_i \tilde{a}(j-i)|x_j|^2 \\
&\leq \sum_j M\gamma_j |x_j|^2 = M\|x\|_\gamma^2.
\end{aligned}
\tag{3.2.21}
$$

The result now follows from Gronwall's lemma. ∎

3.3 Proofs of Theorem 3.1.3 and Lemma 3.1.4

3.3.1 Proof of Lemma 3.1.4

Note that, since any solution X to (3.1.2) solves the martingale problem for the operator A in (3.1.10), we have for any $f \in \mathcal{C}^2_{\text{fin}}(K^\Lambda)$

$$
E[f(X(t))] - E[f(X(0))] = \int_0^t E[Af(X(s))]ds. \tag{3.3.1}
$$

Using the continuity of Af, the continuity of the sample paths of X, and bounded convergence, we see that the function $t \mapsto E[Af(X(t))]$ is continuous. It follows that the function $t \mapsto E[f(X(t))]$ is continuously differentiable and satisfies

$$
\tfrac{\partial}{\partial t} E[f(X(t))] = E[Af(X(t))]. \tag{3.3.2}
$$

Applying the remarks above to the function $f(x) = x_i^\alpha$ and using bounded convergence to interchange an infinite sum and expectation, we see that

$$
\tfrac{\partial}{\partial t} E[X_i^\alpha(t)] = \sum_j a(j-i)(E[X_j^\alpha] - E[X_i^\alpha]). \tag{3.3.3}
$$

When X is shift-invariant, there clearly exist functions $\theta : [0,\infty) \to K$ and $C : [0,\infty) \times \Lambda \to \mathbb{R}$ such that

$$
\left.\begin{aligned}
E[X_i^\alpha(t)] &= \theta^\alpha(t) \\
\mathrm{Cov}(X_i(t), X_j(t)) &= C_t(j-i)
\end{aligned}\right\} \quad (t \geq 0,\ i,j \in \Lambda,\ \alpha = 1,\ldots,d). \tag{3.3.4}
$$

Applying this to (3.3.3), we see that $\frac{\partial}{\partial t}\theta(t) = 0$ and hence

$$
E[X_i(t)] = \theta \qquad (t \geq 0,\ i \in \Lambda) \tag{3.3.5}
$$

for some $\theta \in K$.

Let us put $\tilde{X}_i := X_i - \theta$. Applying (3.3.2) to the function $f(x) = \sum_\alpha (x_i^\alpha - \theta^\alpha)(x_j^\alpha - \theta^\alpha)$, using bounded convergence to interchange an infinite sum and expectation, we get

$$\begin{aligned} &\tfrac{\partial}{\partial t}\mathrm{Cov}(X_i(t), X_j(t)) \\ &\quad = \sum_{k,l} a(k-l) E\Big[\sum_\alpha (\tilde{X}_k^\alpha(t) - \tilde{X}_l^\alpha(t))(\delta_{il}\tilde{X}_j^\alpha(t) + \delta_{jl}\tilde{X}_i^\alpha(t))\Big] \\ &\quad +2\delta_{ij} E[tr(w)(X(t))]. \end{aligned} \tag{3.3.6}$$

Inserting (3.3.4) we get

$$\begin{aligned} \tfrac{\partial}{\partial t} C_t(j-i) = &\sum_k a(k-i)(C_t(j-k) - C_t(j-i)) \\ &+\sum_k a(k-j)(C_t(k-i) - C_t(j-i)) \\ &+2\delta_{ij} E[tr(w)(X(t))]. \end{aligned} \tag{3.3.7}$$

Substituting $\tilde{\imath} := j-i$, $\tilde{\jmath} := k-i$ and $\tilde{k} := j-k$ and reordering the summations, we find that

$$\begin{aligned} \tfrac{\partial}{\partial t} C_t(\tilde{\imath}) = &\sum_{\tilde{\jmath}} a(\tilde{\jmath})(C_t(\tilde{\imath}-\tilde{\jmath}) - C_t(\tilde{\imath})) \\ &+\sum_{\tilde{k}} a(-\tilde{k})(C_t(\tilde{\imath}-\tilde{k}) - C_t(\tilde{\imath})) \\ &+2\delta_{\tilde{\imath}0} E[tr(w)(X(t))]. \end{aligned} \tag{3.3.8}$$

This shows that formula (3.1.34) holds. ∎

3.3.2 Random walk representations

Let $B(\Lambda)$ be the Banach space of bounded real functions on Λ, equipped with the supremum norm. The operator G in (3.1.35) is a bounded linear operator on $B(\Lambda)$. We define a Feller semigroup on $B(\Lambda)$ by

$$P_t f := e^{tG} f, \tag{3.3.9}$$

where $e^{tG} := \sum_{n=0}^\infty \frac{1}{n!}(tG)^n$. This semigroup corresponds to a continuous-time random walk $(I_t)_{t\geq 0}$ on Λ that jumps from i to j with rate $a_S(j-i)$. By shift-invariance there exists a function $P : [0,\infty) \times \Lambda \to \mathbb{R}$ such that

$$P_t(j-i) = P^i[I_t = j]. \tag{3.3.10}$$

We can consider $P_t(j-i)$ as the (i,j)-th element of the matrix of the operator P_t in (3.3.9), in the following sense

$$(P_t f)(i) = \sum_j P_t(j-i) f(j). \tag{3.3.11}$$

Lemma 3.3.1 *Assume that $f, g : [0,\infty) \to B(\Lambda)$ are continuous functions, where $t \mapsto f_t(i)$ is continuously differentiable for each $i \in \Lambda$ and*

$$\tfrac{\partial}{\partial t} f_t(i) = \sum_j a_S(j-i)(f_t(j) - f_t(i)) + g_t(i) \qquad (t \geq 0,\ i \in \Lambda). \tag{3.3.12}$$

Then

$$f_t(i) = \sum_j P_t(j-i) f_0(j) + \int_0^t \sum_j P_s(j-i) g_{t-s}(j) ds \quad (t \geq 0,\ i \in \Lambda). \tag{3.3.13}$$

Proof of Lemma 3.3.1: We define derivatives and Riemann integrals of $B(\Lambda)$-valued functions as in [16], chapter 1. In that language, we would like to rewrite (3.3.12) as

$$\tfrac{\partial}{\partial t} f_t = G f_t + g_t \qquad (t \geq 0). \tag{3.3.14}$$

However, care is needed because it is not immediately clear that the derivative $\frac{\partial}{\partial t} f_t := \lim_{\varepsilon \to 0} \varepsilon^{-1}(f_{t+\varepsilon} - f_t)$ exists in the topology on $B(\Lambda)$. To see that this is all right, we note that the function

$$t \mapsto G f_t + g_t \tag{3.3.15}$$

is continuous in t and therefore

$$t \mapsto \int_0^t (G f_s + g_s) ds \tag{3.3.16}$$

exists and is a continuously differentiable $B(\Lambda)$-valued function. Formula (3.3.12) implies that

$$f_t = \int_0^t (G f_s + g_s) ds \tag{3.3.17}$$

and it follows that $t \mapsto f_t$ is continuously differentiable and (3.3.14) holds. Let $(I_t)_{t \geq 0}$ be the continuous-time random walk with kernel a_S. This process solves the martingale problem for G, and therefore

$$\begin{aligned} E^i[f_t(I_0)] &= E^i[f_0(I_t)] - \int_0^t E^i[((\tfrac{\partial}{\partial s} + G) f_{t-s})(I_s)] ds \\ &= E^i[f_0(I_t)] + \int_0^t E^i[g_{t-s}(I_s)] ds. \end{aligned} \tag{3.3.18}$$

This is formula (3.3.13). ∎

3.3.3 Spatially ergodic measures

The σ-field of shift-invariant events is

$$\mathcal{S} := \{A \in \mathcal{B}(K^{\Lambda}) : T_i^{-1}(A) = A \ \ \forall i \in \Lambda\}. \tag{3.3.19}$$

A probability measure μ on K^{Λ} is spatially ergodic if for every $A \in \mathcal{S}$ either $\mu(A) = 1$ or $\mu(A) = 0$. We state the following standard ergodic theorem in L^2 without proof (see [24]).

Lemma 3.3.2 *For $n = 1, 2, \ldots$, let $p_n : \Lambda \to [0, \infty)$ be functions satisfying $\sum_i p_n(i) = 1$ and*

$$\lim_{n\to\infty} \sum_k |p_n(i-k) - p_n(j-k)| = 0 \qquad \forall i, j \in \Lambda. \tag{3.3.20}$$

Let $X = (X_i)_{i\in\Lambda}$ be a family of K-valued random variables with shift-invariant ergodic law $\mathcal{L}(X)$. If $E[X_0] = \theta$, then

$$\lim_{n\to\infty} E\Big[\Big|\theta - \sum_i p_n(i)X_i\Big|^2\Big] = 0. \tag{3.3.21}$$

In our case, probability distributions p_n satisfying (3.3.20) will arise in the following way.

Lemma 3.3.3 *Let $P : [0, \infty) \times \Lambda \to \mathbb{R}$ be as in (3.3.10). Then for any $i, j \in \Lambda$:*

$$\lim_{t\to\infty} \sum_k |P_t(i-k) - P_t(j-k)| = 0. \tag{3.3.22}$$

Proof of Lemma 3.3.3: We use the Ornstein coupling [25]. To see how this works for random walks on arbitrary Abelian groups, let $\Delta \subset \Lambda$ be a set such that $a_S(k) > 0$ for each $k \in \Delta$ and such that of each $k \in \Lambda$ with $a_S(k) > 0$, either k or $-k$ (but not both) is in Δ. By irreducibility, we can decompose $j - i$ as

$$j - i = \sum_{k\in\Delta} n(k)k, \tag{3.3.23}$$

where $n(k) \in \mathbb{Z}$ and only a finite number of $n(k)$'s are non-zero. We may couple two random walks starting in points i and j in such a way that they always make a jump of size k or $-k$ at the same time. They choose k or $-k$ independently of each other, until the walk starting in j has made $n(k)$ more of these jumps than the walk starting in i. After that, they choose either both k or both $-k$. This coupling is obviously successful and Lemma 3.3.3 now follows easily. ∎

3.3.4 Proof of Theorem 3.1.3

The proof consists of several steps.

$X(\infty)$ **is an invariant law:** By this we mean that there exists a shift-invariant solution X^∞ to the martingale problem for the operator A in (3.1.10) such that

$$\mathcal{L}(X^\infty(t)) = X(\infty) \qquad \forall t \geq 0. \tag{3.3.24}$$

To see this, define solutions to the martingale problem for A by

$$X_n(t) := X(t_n + t), \tag{3.3.25}$$

where (t_n) is some sequence tending to infinity. By Lemma 3.2.2 we can find a subsequence $(X_{n(k)})$ that converges weakly to some solution X^∞ to the martingale problem for A. Now

$$\mathcal{L}(X^\infty(t)) = \lim_{n\to\infty} \mathcal{L}(X(t_n + t)) = \mathcal{L}(X(\infty)) \qquad \forall t \geq 0, \tag{3.3.26}$$

where the limit denotes weak convergence of probability measures on K^Λ. It is easy to see that X^∞ is shift-invariant.

Recurrent a_S**,** $P[X_i(\infty) \in \partial_w K \ \ \forall i \in \Lambda] = 1$**:** Let us write

$$\mathrm{Cov}(X_i^\infty(t), X_j^\infty(t)) = C_t^\infty(j - i) \tag{3.3.27}$$

for covariances belonging to the process X^∞ constructed above. We can apply Lemma 3.1.4 to this process. Lemma 3.3.1 now leads to the representation

$$C_t^\infty(i) - \sum_j P_t(j-i) C_0^\infty(j) = 2\int_0^t P_s(0-i) E[tr(w)(X_0^\infty(t-s))]ds. \tag{3.3.28}$$

By the compactness of the state space K, the left-hand side of (3.3.28) is bounded. The right-hand side is equal to

$$2E[tr(w)(X_0(\infty))] \int_0^t P_s(0-i)ds. \tag{3.3.29}$$

By the recurrence of the random walk with kernel a_S, the integral in (3.3.29) diverges as t tends to infinity, and therefore (3.3.28) can only hold if

$$E[tr(w)(X_0(\infty))] = 0. \tag{3.3.30}$$

This proves that $P[X_0(\infty) \in \partial_w K] = 1$ and by shift-invariance

$$P[X_i(\infty) \in \partial_w K \ \ \forall i \in \Lambda] = 1. \tag{3.3.31}$$

Recurrent a_S, $P[X_i(\infty) = X_j(\infty) \ \ \forall i, j \in \Lambda] = 1$: Applying Lemma 3.1.4 to the process X^∞, we see that

$$\tfrac{\partial}{\partial t} C_t^\infty(i) = \sum_j a_S(j-i)(C_t^\infty(j) - C_t^\infty(i)) + 2\delta_{i0} E[tr(w)(X_0^\infty(t))]. \tag{3.3.32}$$

Here $C_t^\infty(i) = C^\infty(i)$, where we use the notation

$$\mathrm{Cov}(X_i(\infty), X_j(\infty)) = C^\infty(j-i). \tag{3.3.33}$$

Note that $\frac{\partial}{\partial t} C_t^\infty(i) = 0$, while $E[tr(w)(X_0^\infty(t))] = 0$ by (3.3.30). Inserting this into (3.3.32), we get

$$\sum_j a_S(j-i)(C^\infty(j) - C^\infty(i)) = 0. \tag{3.3.34}$$

This means that C^∞ is a bounded a_S-harmonic function. By the Choquet-Deny theorem (which follows easily from Lemma 3.3.3 –see [25], Chapter II, Theorem 1.5) it follows that C^∞ is constant. We write $\tilde{X}_i(t) := X_i(t) - \theta$ with θ as in Lemma 3.1.4 and note that by Cauchy-Schwarz

$$\begin{aligned} C^\infty(j-i) &= E[\tilde{X}_i(\infty) \cdot \tilde{X}_j(\infty)] \\ &\leq E[|\tilde{X}_i(\infty)|^2]^{\frac{1}{2}} E[|\tilde{X}_j(\infty)|^2]^{\frac{1}{2}} = E[|\tilde{X}_0(\infty)|^2] = C^\infty(0), \end{aligned} \tag{3.3.35}$$

where equality holds if and only if $P[X_i(\infty) = X_j(\infty)] = 1$. This proves that

$$P[X_i(\infty) = X_j(\infty) \ \ \forall i, j \in \Lambda] = 1. \tag{3.3.36}$$

Transient a_S, $P[X_i(\infty) \in \partial_w K] < 1 \ \forall i \in \Lambda$: We start by noting that the ergodicity of $\mathcal{L}(X(0))$ implies that for each $i \in \Lambda$

$$\lim_{t\to\infty} \sum_j P_t(j-i) C_0(j) = 0. \tag{3.3.37}$$

To see this, write $\tilde{X}_i(0) := X_i(0) - \theta$ as before and note that by Lemma 3.3.2 and 3.3.3

$$\lim_{t\to\infty} E\Big[\Big|\sum_j P_t(j)\tilde{X}_j(0)\Big|^2\Big] = 0. \tag{3.3.38}$$

Here

$$
\begin{aligned}
E\Big[\Big|\sum_j P_t(j)\tilde{X}_j(0)\Big|^2\Big] &= \sum_{jk} P_t(j)P_t(k)E[\tilde{X}_j(0)\tilde{X}_k(0)] \\
&= \sum_{jk} P_t(j)P_t(k)C_0(k-j) \\
&= \sum_{ij} P_t(j)P_t(i+j)C_0(i) \\
&= \sum_i \Big(\sum_j P_t(j)P_t(i-j)\Big)C_0(i) \\
&= \sum_i P_{2t}(i)C_0(i),
\end{aligned}
\tag{3.3.39}
$$

where all infinite sums are absolutely convergent and we have used that, by the symmetry of a_S, $P_t(i) = P_t(-i)$. Formula (3.3.38) and (3.3.39) show that (3.3.37) holds for $i = 0$. Using Lemma 3.3.3 we can easily generalize this to arbitrary $i \in \Lambda$.

By Lemma 3.1.4 and Lemma 3.3.1 we have the representation

$$
C_t(i) = \sum_j P_t(j-i)C_0(j) + 2\int_0^t P_s(0-i)E[tr(w)(X_0(t-s))]ds. \tag{3.3.40}
$$

Taking the limit $t \to \infty$ we get with the help of (3.3.37) that

$$
\begin{aligned}
C^\infty(i) &= \lim_{t\to\infty} 2\int_0^t P_s(0-i)E[tr(w)(X_0(t-s))]ds \\
&= 2E[tr(w)(X_0(\infty))]\int_0^\infty P_t(0-i)dt,
\end{aligned}
\tag{3.3.41}
$$

where we use the notation in (3.3.33). Let us assume for the moment that $E[tr(w)(X_0(\infty))] = 0$. Then $P[X_0(\infty) \in \partial_w K] = 1$. On the other hand, (3.3.41) gives $C^\infty(0) = 0$ and hence $P[X_0(\infty) = \theta] = 1$. This contradicts our assumption that $\theta \notin \partial_w K$ and we conclude that $E[tr(w)(X_0(\infty))] > 0$. Therefore $P[X_0(\infty) \in \partial_w K] < 1$ and the claim follows from shift-invariance.

Transient a_S, $P[X_i(\infty) = X_j(\infty)] < 1\ \forall i \neq j \in \Lambda$: Let $(I_t)_{t\geq 0}$ be the random walk with kernel a_S. Let τ_i be the stopping time

$$
\tau_i := \inf\{t \geq 0 : I_t = i\} \qquad (i \in \Lambda). \tag{3.3.42}
$$

It is easy to see that for all $i \in \Lambda$

$$
\int_0^\infty P_t(0-i)dt = P^i[\tau_0 < \infty]\int_0^\infty P_t(0)dt. \tag{3.3.43}
$$

Let us assume that for some $i \neq 0$ we have $P^i[\tau_0 < \infty] = 1$. Then by the symmetry of the random walk, also $P^0[\tau_i < \infty] = 1$. But this implies that the random walk starting in 0 visits 0 infinitely often, which contradicts our assumption that it is transient. It follows that $P^i[\tau_j < \infty] < 1$ for all $i \neq j$. Combining (3.3.43) and (3.3.41) we can conclude that

$$C^{\infty}(i) < C^{\infty}(0) \qquad \forall i \neq 0. \tag{3.3.44}$$

Now Cauchy-Schwarz in (3.3.35) implies that $P[X_i(\infty) = X_j(\infty)] < 1$ for all $i \neq j$. ∎

3.4 Proofs of Theorem 3.1.5, Lemma 3.1.6 and Corollary 3.1.7

3.4.1 Potential theory

In this section we collect some elementary facts about w-harmonic functions from potential theory. We assume that for each $x \in K$, the non-interacting equation (3.1.19) has a unique weak solution X^x with initial condition $X^x(0) = x$. We denote its last element by $X^x(\infty)$. We denote the semigroup on $B(K)$ associated with (3.1.19) by $(S_t)_{t\geq 0}$ and we add a last element S_∞ as in (3.1.41). Restricted to the smaller domain $\mathcal{C}(K) \subset B(K)$, the $(S_t)_{t\geq 0}$ form a Feller semigroup, whose generator we denote by G. Note that $\mathcal{D}(G) \subset \mathcal{C}(K)$.

Lemma 3.4.1 *For each solution X to (3.1.19)*

$$P[X(\infty) \in \partial_w K] = 1. \tag{3.4.1}$$

Proof of Lemma 3.4.1: Since X is a bounded martingale, it converges. Now the lemma is just a special case of Theorem 3.1.3. ∎

Lemma 3.4.2 *Assume that $S_\infty(\mathcal{C}(K)) \subset \mathcal{C}(K)$. Consider sets H, H', H'', H''' defined as*

$$\begin{aligned}
H &:= \{f \in \mathcal{D}(G) : Gf = 0\} \\
H' &:= \{f \in \mathcal{C}(K) : S_t f = f \ \ \forall t \in [0,\infty]\} \\
H'' &:= \{f \in \mathcal{C}(K) : S_\infty f = f\} \\
H''' &:= \{S_\infty \phi : \phi \in \mathcal{C}(K)\}
\end{aligned} \tag{3.4.2}$$

Then $H = H' = H'' = H'''$. For each $\phi \in \mathcal{C}(K)$ there exists a unique $f \in H$ such that

$$f(x) = \phi(x) \qquad (x \in \partial_w K) \tag{3.4.3}$$

and this f is given by

$$f = S_\infty \phi. \tag{3.4.4}$$

Proof of Lemma 3.4.2: It is easy to see that $H \subset H' \subset H'' \subset H'''$. To see that $H''' \subset H$, note that $\phi(X(t)) \to \phi(X(\infty))$ almost surely, so bounded convergence implies that $E^x[\phi(X(t))] \to E^x[\phi(X(\infty))]$ for each $x \in K$. Since $|E^x[\phi(X(t))]| \leq \|\phi\|_\infty < \infty$, it follows that $S_t\phi \to S_\infty\phi$ as $t \to \infty$ in the sense of bounded pointwise convergence. Therefore

$$(S_t S_\infty \phi)(x) = \lim_{s\to\infty} (S_t S_s \phi)(x) = (S_\infty \phi)(x) \quad (x \in K). \tag{3.4.5}$$

It follows that $t^{-1}(S_t - 1)S_\infty\phi = 0$ for all t, so

$$\lim_{t\to 0} t^{-1}(S_t - 1)S_\infty \phi = 0 \tag{3.4.6}$$

in the topology on $\mathcal{C}(K)$ and this proves that $H''' \subset H$.

By (3.4.2), $S_\infty\phi \in H$ for each $\phi \in \mathcal{C}(K)$. To see that $f := S_\infty\phi$ solves (3.4.3) it suffices to note that for each $x \in \partial_w K$ the process

$$X(t) := x \tag{3.4.7}$$

solves (3.1.19). To see that f is the unique w-harmonic function satisfying (3.4.3), suppose that $\tilde{f} \in H$ is another one. Then by (3.4.2) and by Lemma 3.4.1

$$\tilde{f} = S_\infty \tilde{f} = S_\infty \phi = f. \tag{3.4.8}$$

∎

In the proof of Theorem 3.1.5 we will make use of the function v^*, given by

$$v^*(x) = tr(w^*), \tag{3.4.9}$$

where w^* is the special diffusion matrix mentioned in (3.1.52). The following lemma collects some elementary facts about v^*.

Lemma 3.4.3 *Assume that $S_\infty(\mathcal{C}(K)) \subset \mathcal{C}(K)$. Then there exists a unique function $v^* \in \mathcal{D}(G)$ such that*

$$\begin{aligned} -\tfrac{1}{2}(Gv^*)(x) &= tr(w)(x) \qquad && (x \in K) \\ v^*(x) &= 0 && (x \in \partial_w K). \end{aligned} \tag{3.4.10}$$

This function v^* satisfies

$$\begin{aligned} v^*(x) &\geq 0 \qquad && x \in K \\ v^*(x) &= 0 \quad \Leftrightarrow && x \in \partial_w K, \end{aligned} \tag{3.4.11}$$

and is given by the formula

$$v^*(x) = \mathrm{Var}(X^x(\infty)). \tag{3.4.12}$$

Proof of Lemma 3.4.3: We write $\underline{x}^2$ for the function $x \mapsto |x|^2$. Then

$$G\underline{x}^2 = 2tr(w). \tag{3.4.13}$$

Thus, (3.4.10) can be rewritten as

$$\begin{aligned} G(v^* + \underline{x}^2) &= 0 \qquad \text{on } K \\ (v^* + \underline{x}^2) &= \underline{x}^2 \qquad \text{on } \partial_w K. \end{aligned} \tag{3.4.14}$$

Lemma 3.4.2 shows that $v^* + \underline{x}^2$ can be uniquely solved from these equations. Using the fact that X^x solves the martingale problem for G we see with the help of (3.4.10) and (3.4.13) that

$$\begin{aligned} E\left|X^x(\infty) - x\right|^2 &= 2\int_0^\infty E[tr(w)(X(t))]dt \\ &= -E[v^*(X^x(\infty))] + v^*(x). \end{aligned} \tag{3.4.15}$$

Lemma 3.4.1 implies that $E[v^*(X^x(\infty))] = 0$ and so we see that (3.4.12) holds. Formula (3.4.12) immediately implies that $v^*(x) \geq 0$. Finally $v^*(x) = 0$ implies $\mathrm{Var}(X^x(\infty)) = 0$ so that $X^x(\infty) = x$, and by Lemma 3.4.1 this in turn implies that $x \in \partial_w K$. ∎

3.4.2 Infinite-dimensional differentiation

We will need to extend the domain of the operator A in (3.1.10) to include functions depending on infinitely many coordinates. In order to do this properly, we introduce the space $\mathcal{C}^2_{\text{sum}}(K^\Lambda)$ of functions with summable continuous second derivatives. For a function $f : (\mathbb{R}^d)^\Lambda \to \mathbb{R}$ we define $\frac{\partial}{\partial x_i^\alpha} f(x)$ in the usual way. $\mathcal{C}^2((\mathbb{R}^d)^\Lambda)$ is the class of functions for which all zeroth, first and second order derivatives are continuous functions on $(\mathbb{R}^d)^\Lambda$. $\mathcal{C}^2(K^\Lambda)$ is the set of funtions on K^Λ that can be extended to functions in $\mathcal{C}^2(\mathbb{R}^d)$. $\mathcal{C}^2_{\text{sum}}(K^\Lambda)$, finally, is the space of functions in $\mathcal{C}^2(K^\Lambda)$ for which

$$\begin{aligned} x &\mapsto \left(\frac{\partial}{\partial x_i^\alpha} f(x)\right)_{i\in\Lambda}^{\alpha=1,\dots,d} \\ x &\mapsto \left(\frac{\partial^2}{\partial x_i^\alpha \partial x_j^\beta} f(x)\right)_{i,j\in\Lambda}^{\alpha,\beta=1,\dots,d} \end{aligned} \tag{3.4.16}$$

are continuous maps from K^Λ into $l^1(\{1,\dots,d\} \times \Lambda)$ and $l^1(\{1,\dots,d\}^2 \times \Lambda^2)$, respectively, spaces of absolutely summable sequences, equipped with the l^1-norm.

Lemma 3.4.4 *For $i, j \in \Lambda$ and $\alpha, \beta = 1, \ldots, d$, let $b_{i,\alpha}$ and $a_{ij,\alpha\beta}$ be functions in $\mathcal{C}(K^\Lambda)$ satisfying uniform bounds*

$$\begin{aligned} &\|b_{i,\alpha}\|_\infty \le M_1 && \forall i \in \Lambda,\ \alpha = 1, \ldots, d \\ &\|a_{ij,\alpha\beta}\|_\infty \le M_2 && \forall i, j \in \Lambda,\ \alpha, \beta = 1, \ldots, d. \end{aligned} \tag{3.4.17}$$

Then for each $f \in \mathcal{C}^2_{\text{sum}}(K^\Lambda)$ and for all finite $\Delta_n \subset \Lambda$ with $\Delta_n \uparrow \Lambda$, the limit

$$\lim_{n\to\infty} \Big(\sum_{i\in\Delta_n,\ \alpha} b_{i,\alpha}(x) \tfrac{\partial}{\partial x_i^\alpha} + \sum_{ij\in\Delta_n,\ \alpha\beta} a_{ij,\alpha\beta}(x) \tfrac{\partial^2}{\partial x_i^\alpha \partial x_j^\beta} \Big) f(x) \tag{3.4.18}$$

exists in the topology on $\mathcal{C}(K^\Lambda)$ and does not depend on the choice of the Δ_n.

Proof of Lemma 3.4.4: We treat only the convergence of the first order derivatives; the argument for second order derivatives is then the same. Define operators (A_n) by

$$(A_n f)(x) := \sum_{i\in\Delta_n,\alpha} b_{i,\alpha}(x) \tfrac{\partial}{\partial x_i^\alpha} f(x). \tag{3.4.19}$$

For each $n \le m$ and for each $f \in \mathcal{C}^2_{\text{sum}}(K^\Lambda)$ we have

$$\begin{aligned} \|A_n f - A_m f\|_\infty &= \sup_{x\in K^\Lambda} \Big| \sum_{i\in\Delta_m\setminus\Delta_n,\ \alpha} b_{i,\alpha}(x) \tfrac{\partial}{\partial x_i^\alpha} f(x) \Big| \\ &\le M_1 \sup_{x\in K^\Lambda} \sum_{i\in\Lambda\setminus\Delta_n,\ \alpha} \Big| \tfrac{\partial}{\partial x_i^\alpha} f(x) \Big|. \end{aligned} \tag{3.4.20}$$

The functions $g_n : K \to l^1(\{1, \ldots, d\} \times \Lambda)$ given by

$$g_n(x) := \sum_{i\in\Lambda\setminus\Delta_n,\ \alpha} \Big| \tfrac{\partial}{\partial x_i^\alpha} f(x) \Big| \tag{3.4.21}$$

are continuous functions decreasing to zero as $n \to \infty$, and hence by Dini's theorem $g_n \to 0$ uniformly. Thus we see by (3.4.20) that the sequence $(A_n f)$ is a Cauchy sequence in the Banach space $\mathcal{C}(K^\Lambda)$.

To see that the limit does not depend on the summation order, observe that for each $x \in K^\Lambda$

$$\sum_{i,\alpha} \Big| b_{i,\alpha}(x) \tfrac{\partial}{\partial x_i^\alpha} f(x) \Big| \le M_1 \sum_{i,\alpha} \Big| \tfrac{\partial}{\partial x_i^\alpha} f(x) \Big| < \infty. \tag{3.4.22}$$

This means that we can write

$$\lim_{n\to\infty} (A_n f)(x) = \sum_{i,\alpha} b_{i,\alpha}(x) \tfrac{\partial}{\partial x_i^\alpha} f(x), \tag{3.4.23}$$

where the sums are pointwise absolutely convergent so that the result does not depend on the summation order. ∎

For each $i \in \Lambda$ and $\alpha, \beta \in \{1, \ldots, d\}$ the maps

$$\begin{aligned} x &\mapsto \sum_j a(j-i)(x_j^\alpha - x_i^\alpha) \\ x &\mapsto w_{\alpha\beta}(x_i) \end{aligned} \tag{3.4.24}$$

are continuous on K^Λ and satisfy the uniform bounds

$$\begin{aligned} &\sup_{x \in K^\Lambda} \Big| \sum_j a(j-i)(x_j^\alpha - x_i^\alpha) \Big| \leq \Big(\sum_k a(k) \Big) \sup_{\alpha'} \sup_{y,z \in K} \big| y^{\alpha'} - z^{\alpha'} \big| \\ &\sup_{x \in K^\Lambda} \big| w_{\alpha\beta}(x_i) \big| \leq \sup_{\alpha'\beta'} \| w_{\alpha'\beta'} \|_\infty. \end{aligned} \tag{3.4.25}$$

Therefore, by Lemma 3.4.4, the operator

$$(A'f)(x) := \Big(\sum_{i,\alpha} \sum_j a(j-i)(x_j^\alpha - x_i^\alpha) \tfrac{\partial}{\partial x_i^\alpha} + \sum_{i,\alpha\beta} w_{\alpha\beta}(x_i) \tfrac{\partial^2}{\partial x_i^\alpha \partial x_i^\beta} \Big) f(x) \tag{3.4.26}$$

is well-defined on $\mathcal{C}^2_{\rm sum}(K^\Lambda)$, where the infinite sums are convergent in $\mathcal{C}(K^\Lambda)$ and the result is independent of the summation order. We now show that A is a core for A'.

Lemma 3.4.5 *Let $\overline{A}$ be the closure of the operator A in (3.1.10), and let $\mathcal{D}(\overline{A})$ be its domain. Then $\mathcal{C}^2_{\rm sum}(K^\Lambda) \subset \mathcal{D}(\overline{A})$ and*

$$(\overline{A}f)(x) = \Big(\sum_{i,\alpha} \sum_j a(j-i)(x_j^\alpha - x_i^\alpha) \tfrac{\partial}{\partial x_i^\alpha} + \sum_{i,\alpha\beta} w_{\alpha\beta}(x_i) \tfrac{\partial^2}{\partial x_i^\alpha \partial x_i^\beta} \Big) f(x) \tag{3.4.27}$$

for each $f \in \mathcal{C}^2_{\rm sum}(K^\Lambda)$.

Proof of Lemma 3.4.5: We have to show that for each $f \in \mathcal{C}^2_{\rm sum}(K^\Lambda)$ there exist $f_n \in \mathcal{D}(A)$ such that $f_n \to f$ and $Af_n \to A'f$, with A' as in (3.4.26). Fix $f \in \mathcal{C}^2_{\rm sum}(K^\Lambda)$ and define π_n, f_n as in (3.2.1) and (3.2.2), so that $f_n \to f$. It is easy to see that there exists a constant C such that

$$\begin{aligned} &\| Af_n(x) - A'f(x) \|_\infty \\ &\quad \leq C \sup_{x \in K^\Lambda} \Bigg(\sum_{i,\alpha} \Big| \tfrac{\partial}{\partial x_i^\alpha} f_n(x) - \tfrac{\partial}{\partial x_i^\alpha} f(x) \Big| + \sum_{i,\alpha\beta} \Big| \tfrac{\partial^2}{\partial x_i^\alpha \partial x_i^\beta} f_n(x) - \tfrac{\partial^2}{\partial x_i^\alpha \partial x_i^\beta} f(x) \Big| \Bigg) \\ &\quad \leq C \sup_{x \in K^\Lambda} \Bigg(\sum_{i \leq n,\alpha} \Big| (\tfrac{\partial}{\partial x_i^\alpha} f)(\pi_n(x)) - \tfrac{\partial}{\partial x_i^\alpha} f(x) \Big| + \sum_{i > n,\alpha} \Big| \tfrac{\partial}{\partial x_i^\alpha} f(x) \Big| \\ &\qquad\qquad + \sum_{i \leq n,\alpha\beta} \Big| (\tfrac{\partial^2}{\partial x_i^\alpha \partial x_i^\beta} f)(\pi_n(x)) - \tfrac{\partial^2}{\partial x_i^\alpha \partial x_i^\beta} f(x) \Big| + \sum_{i > n,\alpha\beta} \Big| \tfrac{\partial^2}{\partial x_i^\alpha \partial x_i^\beta} f(x) \Big| \Bigg). \end{aligned} \tag{3.4.28}$$

By the compactness of K^Λ, the maps

$$\begin{aligned} x &\mapsto \left(\tfrac{\partial}{\partial x_i^\alpha} f(x)\right)_{i\in\Lambda}^{\alpha=1,\ldots,d} \\ x &\mapsto \left(\tfrac{\partial^2}{\partial x_i^\alpha \partial x_j^\beta} f(x)\right)_{i,j\in\Lambda}^{\alpha,\beta=1,\ldots,d} \end{aligned} \tag{3.4.29}$$

are uniformly continous with respect to the norm on the spaces $l^1(\{1,\ldots,d\}\times\Lambda)$ and $l^1(\{1,\ldots,d\}^2\times\Lambda^2)$. This implies that

$$\lim_{n\to\infty}\sup_{x\in K^\Lambda}\sum_{i,\alpha}\left|\left(\tfrac{\partial}{\partial x_i^\alpha}f\right)(\pi_n(x)) - \tfrac{\partial}{\partial x_i^\alpha}f(x)\right| = 0, \tag{3.4.30}$$

and similarly for second derivatives. Finally, by Dini's theorem (see (3.4.21))

$$\lim_{n\to\infty}\sup_{x\in K^\Lambda}\sum_{i>n,\alpha}\left|\tfrac{\partial}{\partial x_i^\alpha}f(x)\right| = 0, \tag{3.4.31}$$

and similarly for second derivatives, so $Af_n \to A'f$. ∎

3.4.3 Proof of Lemma 3.1.6

The model with zero diffusion: In the special case where $w = 0$, the system of stochastic differential equations (3.1.2) reduces to

$$dX_i(t) = \sum_{j\in\Lambda} a(j-i)(X_j(t) - X_i(t))dt \qquad (i\in\Lambda,\ t\geq 0). \tag{3.4.32}$$

By Theorems 3.1.1 and 3.1.2 this system of equations has a unique solution. We can write down the solution of (3.4.32) explicitly in terms of the random walk on Λ that jumps from i to j with rate $a(j-i)$. Let $P_t(j-i)$ denote the probability that this random walk, starting in i at time 0, is in j at time t. Then the unique solution of (3.4.32) is given by (see Lemma 3.3.1)

$$X_i(t) = \sum_j P_t(j-i)X_j(0). \tag{3.4.33}$$

Let $(P_t)_{t\geq 0}$ be the semigroup on $B(\Lambda)$ associated with the random walk with kernel a (see section 3.3.2). Let us denote by $(R_t)_{t\geq 0}$ the Feller semigroup on $\mathcal{C}(K^\Lambda)$ associated with the process in (3.4.32):

$$(R_t f)(x) := E[f(X^x(t))] = f(P_t x). \tag{3.4.34}$$

Applying Lemma 3.4.5 to the case $w = 0$, we see that the generator of $(R_t)_{t\geq 0}$ is an extension of the operator

$$(Bf)(x) := \sum_{i,\alpha}\sum_j a(j-i)(x_j - x_i)\tfrac{\partial}{\partial x_i^\alpha} f(x) \tag{3.4.35}$$

with domain $\mathcal{D}(B) := \mathcal{C}^2_{\text{sum}}(K^\Lambda)$.

Evolution of harmonic functions: We now set out to prove Lemma 3.1.6. We start with the case $f \in \mathcal{C}^2(K) \cap H$. Fix $i \in \Lambda$, and let $h \in \mathcal{C}^2_{\text{fin}}(K^\Lambda)$ be given by

$$h(x) := f(x_i). \tag{3.4.36}$$

In the language above, we want to show that

$$E[h(X(t))] = E[(R_t h)(X(0))]. \tag{3.4.37}$$

It is not hard to see that $R_t h \in \mathcal{C}^2_{\text{sum}}(K^\Lambda)$ for each $t \geq 0$, where by (3.4.34)

$$\begin{aligned}\tfrac{\partial}{\partial x^\alpha_k}(R_t h)(x) &= P_t(k-i)\Big(\tfrac{\partial}{\partial x^\alpha} f\Big)\Big(\sum_j P_t(j-i)x_j\Big)\\ \tfrac{\partial^2}{\partial x^\alpha_k \partial x^\beta_l}(R_t h)(x) &= P_t(k-i)P_t(l-i)\Big(\tfrac{\partial^2}{\partial x^\alpha \partial x^\beta} f\Big)\Big(\sum_j P_t(j-i)x_j\Big).\end{aligned} \tag{3.4.38}$$

General theory (see [16], chapter 1) now tells us that $t \mapsto R_t h$ is continuously differentiable in $\mathcal{C}(K)$ and

$$\tfrac{\partial}{\partial t} R_t h = B R_t h, \tag{3.4.39}$$

with B the operator in (3.4.35). By Lemma 3.4.5, X solves the martingale problem for the operator

$$A' := B + C, \tag{3.4.40}$$

where

$$(Cf)(x) := \sum_{i,\alpha\beta} w_{\alpha\beta}(x_i) \tfrac{\partial^2}{\partial x^\alpha_i \partial x^\beta_i} f(x) \tag{3.4.41}$$

and $\mathcal{D}(A') = \mathcal{C}^2_{\text{sum}}(K^\Lambda)$. It follows that

$$\begin{aligned} E[(R_0 h)(X(T))] - E[(R_T h)(X(0))] &= E\int_0^T (B + C + \tfrac{\partial}{\partial t})(R_{T-t}h)(X(t))dt\\ &= E\int_0^T (C R_{T-t}h)(X(t))dt.\end{aligned} \tag{3.4.42}$$

By (3.4.38) we have, for any $x \in K^\Lambda$

$$(C R_{T-t}h)(x) = \sum_j P_{T-t}(j-i)^2 \sum_{\alpha\beta} w_{\alpha\beta}(x_j)\Big(\tfrac{\partial^2}{\partial x^\alpha \partial x^\beta} f\Big)\Big(\sum_j P_t(j-i)x_j\Big). \tag{3.4.43}$$

Using Lemma 3.4.2 it is not hard to see that the stability of the boundary distribution against a linear drift is equivalent to formula (3.1.46). The semigroup $(T_{\theta,t})_{t\geq 0}$ maps differentiable functions into differentiable functions, and hence

$$T_{\theta,t}(\mathcal{C}^2(K)\cap H)\subset \mathcal{C}^2(K)\cap H \qquad \forall\theta\in K,\ t\geq 0. \tag{3.4.44}$$

This means that $GT_{\theta,t}f=0$ for all $f\in\mathcal{C}^2(K)\cap H$ and $\theta\in K$, $t\geq 0$, which says that for any $x\in K$

$$\begin{aligned}&\sum_{\alpha\beta} w_{\alpha\beta}(x)\tfrac{\partial^2}{\partial x^\alpha\partial x^\beta}(f(\theta+(x-\theta)e^{-t}))\\ &\qquad = e^{-2t}\sum_{\alpha\beta} w_{\alpha\beta}(x)(\tfrac{\partial^2}{\partial x^\alpha\partial x^\beta}f)(e^{-t}x+(1-e^{-t})\theta)=0 \\ &\qquad\qquad\qquad\qquad\qquad \forall f\in\mathcal{C}^2(K)\cap H,\ \theta\in K,\ t\geq 0.\end{aligned} \tag{3.4.45}$$

For the x here we insert the x_i in (3.4.43) and we fit θ and t such that $e^{-t}=P_t(0)$ and $(1-e^{-t})\theta=\sum_{j\neq i}P_t(j-i)x_j$. Inserting this into (3.4.43) we see that each term in the sum over j there is zero, and therefore (3.4.42) gives

$$E[f(X_i(T))]=E\Big[f\Big(\sum_j P_T(j-i)X_j(0)\Big)\Big]. \tag{3.4.46}$$

To generalize this to arbitrary $f\in H$ it suffices to note that the set of functions $f\in B(K)$ for which (3.4.46) holds is bp-closed. ∎

3.4.4 Proof of Theorem 3.1.5

Comparison argument: The function $x\mapsto tr(w)(x)$ is continuous, takes only non-negative values, and satisfies

$$tr(w)(x)=0\Leftrightarrow x\in\partial_w K. \tag{3.4.47}$$

The same is true for the function $x\mapsto v^*(x)$ (see Lemma 3.4.3) and therefore for each $\varepsilon>0$ we can find a $\lambda>0$ such that

$$tr(w)(x)\geq\lambda(v^*(x)-\varepsilon) \qquad (x\in K). \tag{3.4.48}$$

When we insert this inequality into formula (3.1.34) in Lemma 3.1.4 we see that for all $i\in\Lambda$, $t\geq 0$

$$\tfrac{\partial}{\partial t}C_t(i)\geq\sum_j a_S(j-i)(C_t(j)-C_t(i))+2\lambda\delta_{i0}(E[v^*(X_0(t))]-\varepsilon). \tag{3.4.49}$$

We apply Lemma 3.4.3 to see that the function

$$x \mapsto v^*(x) + |x - \theta|^2 \tag{3.4.50}$$

is w-harmonic. Lemma 3.1.6 therefore tells us that for all $t \geq 0$

$$E[v^*(X_0(t))] + \mathrm{Var}(X_0(t)) = E\Big[v^*\Big(\sum_j P_t(j)X_j(0)\Big)\Big] + \mathrm{Var}\Big(\sum_j P_t(j)X_j(0)\Big). \tag{3.4.51}$$

By Lemma 3.3.3, Lemma 3.3.2 and the spatial ergodicity of $\mathcal{L}(X(0))$ this implies that

$$\lim_{t\to\infty} E[v^*(X_0(t))] + C_t(0) = v^*(\theta). \tag{3.4.52}$$

Combining this with (3.4.49) we see there exists a T such that for all $t \geq T$

$$\tfrac{\partial}{\partial t}C_t(i) \geq \sum_j a_S(j-i)(C_t(j) - C_t(i)) + 2\lambda\delta_{i0}(v^*(\theta) - C_t(0) - 2\varepsilon). \tag{3.4.53}$$

Random walk representation: Let us define

$$D_t(i) := v^*(\theta) - C_t(i) - 2\varepsilon \qquad (i \in \Lambda,\ t \geq 0). \tag{3.4.54}$$

Then (3.4.53) can be rewritten as

$$\tfrac{\partial}{\partial t}D_t(i) \leq \sum_j a_S(j-i)(D_t(j) - D_t(i)) - 2\lambda\delta_{i0}D_t(0) \qquad (t \geq T). \tag{3.4.55}$$

We note that since $t \mapsto C_t$ is continuously differentiable in $B(\Lambda)$, so is $t \mapsto D_t$. Arguing as in the proof of Lemma 3.3.1, we can represent solutions of the differential inequality (3.4.55) in terms of a contracting semigroup $(P^\lambda_t)_{t\geq 0}$ on $B(\Lambda)$, with generator

$$(Gf)(i) := \sum_j a_S(j-i)(f(j) - f(i)) - 2\lambda\delta_{i0}f(0) \qquad (f \in B(\Lambda)). \tag{3.4.56}$$

This semigroup is related to a random walk on Λ that jumps from i to j with rate $a_S(j-i)$ and that is killed at the origin with rate 2λ. When $P^\lambda_t(j,i)$ denotes the probability that this random walk, starting from a point i, is in j at time t, then

$$(P^\lambda_t f)(i) = \sum_j P^\lambda_t(j,i)f(j) \qquad (f \in B(\Lambda)), \tag{3.4.57}$$

and for solutions of (3.4.55) we have the representation

$$D_{T+t}(i) \leq \sum_j P^\lambda_t(j,i)D_T(j) \qquad (t \geq 0). \tag{3.4.58}$$

Convergence of the covariance function: If a_S is recurrent, then the random walk is killed with probability one. This means that for each $i \in \Lambda$

$$\lim_{t\to\infty} \sum_j P_t^\lambda(j,i) = 0. \tag{3.4.59}$$

Combining this with (3.4.58) and using the boundedness of K we see that for each $i \in \Lambda$ there exists a T' such that for all $t \geq T'$

$$C_t(i) \geq v^*(\theta) - 3\varepsilon. \tag{3.4.60}$$

We have thus shown that for every $i \in \Lambda$

$$\liminf_{t\to\infty} C_t(i) \geq v^*(\theta). \tag{3.4.61}$$

On the other hand, with the help of formula (3.4.52) it is easy to see that

$$\limsup_{t\to\infty} C_t(0) \leq v^*(\theta). \tag{3.4.62}$$

By Cauchy-Schwarz we have $C_t(i) \leq C_t(0)$ for all $i \in \Lambda$ (compare (3.3.35)), and hence

$$\lim_{t\to\infty} C_t(i) = v^*(\theta) \qquad \forall i \in \Lambda. \tag{3.4.63}$$

Convergence of $X_0(t)$: Let $(S_t)_{t\geq 0}$ be the semigroup in (3.1.40). Pick any function $\phi \in \mathcal{C}(K)$. By Lemma 3.4.2

$$\phi(x) - (S_\infty \phi)(x) = 0 \qquad (x \in \partial_w K). \tag{3.4.64}$$

Formulas (3.4.52) and (3.4.63) imply that

$$\lim_{t\to\infty} E[v^*(X_0(t))] = 0. \tag{3.4.65}$$

Since v^* is continuous, non-negative, and zero only at $\partial_w K$ (see Lemma 3.4.3) formulas (3.4.64) and (3.4.65) imply that

$$\lim_{t\to\infty} \Big(E[\phi(X_0(t))] - E[(S_\infty\phi)(X_0(t))] \Big) = 0. \tag{3.4.66}$$

Now $S_\infty \phi \in H$ (Lemma 3.4.2) and therefore Lemmas 3.1.6, 3.3.2 and 3.3.3 imply that

$$\lim_{t\to\infty} E[(S_\infty\phi)(X_0(t))] = (S_\infty\phi)(\theta) = \int_K \Gamma_\theta(dx)\phi(x). \tag{3.4.67}$$

Thus we see that

$$X_0(t) \Rightarrow X_0(\infty) \qquad \text{as } t \to \infty, \tag{3.4.68}$$

where the law of $X_0(\infty)$ is given by

$$\mathcal{L}(X_0(\infty)) = \Gamma_\theta. \tag{3.4.69}$$

Convergence of $X(t)$: Formula (3.4.61) and Cauchy-Schwarz imply that for all $i, j \in \Lambda$

$$\lim_{t\to\infty} E|X_i(t) - X_j(t)|^2 = 0. \tag{3.4.70}$$

Combining this with (3.4.68) we easily see that for each finite $\Delta \subset \Lambda$ the collection $(X_i(t))_{i\in\Delta}$ converges weakly to a limit $(X_i(\infty))_{i\in\Delta}$. By the fact that continuous functions depending on finitely many coordinates only are dense in $\mathcal{C}(K^\Lambda)$ (see the proof of Lemma 3.2.1) this implies weak convergence of $X(t)$. ∎

3.4.5 Proof of Corollary 3.1.7

Under the condition $w = \lambda w^*$, most inequalities in the proof of Theorem 3.1.5 can be replaced by equalities. In fact, under the weaker condition (recall that $v^* = tr(w^*)$)

$$tr(w) = \lambda v^*, \tag{3.4.71}$$

we have equality in (3.4.48) with $\varepsilon = 0$. Since we are working with the initial condition $X_i(0) = \theta$ for all $i \in \Lambda$, formula (3.4.52) strengthens to

$$E[v^*(X_0(t))] + C_t(0) = v^*(\theta) \qquad (t \geq 0). \tag{3.4.72}$$

Formula (3.4.58) with $T = 0 = \varepsilon$ and an equality sign reads

$$(v^*(\theta) - C_t(i)) = \sum_j P_t^\lambda(j,i)(v^*(\theta) - C_0(j)), \tag{3.4.73}$$

where $C_0(j) = 0$ for all j. In this way we find that

$$C_t(i) = v^*(\theta)\Big(1 - \sum_j P_t^\lambda(j,i)\Big). \tag{3.4.74}$$

Here $1 - \sum_j P_t^\lambda(j,i)$ is the same as the probability $K_t^\lambda(i)$ appearing in (3.1.54). One can derive formula (3.1.54) in a similar way as formula (3.4.74). For that, one needs to replace the covariance function $C_t(i)$ by a covariance matrix function

$$C_t(j-i)_{\alpha\beta} := E[(X_i^\alpha(t) - \theta^\alpha)(X_j^\beta(t) - \theta^\beta)]. \tag{3.4.75}$$

Generalizing Lemma 3.4.3 one then finds that, for each α, β, the function

$$x \mapsto w^*_{\alpha\beta}(x) \tag{3.4.76}$$

is the unique function in $\mathcal{D}(G)$ solving

$$\begin{aligned} -\tfrac{1}{2}Gw^*_{\alpha\beta}(x) &= w_{\alpha\beta}(x) && (x \in K) \\ w^*_{\alpha\beta}(x) &= 0 && (x \in \partial_w K). \end{aligned} \tag{3.4.77}$$

The rest of the proof is now in complete analogy with the proof of formula (3.4.74). ∎

3.5 Proofs of the examples

3.5.1 Proof of Example 3.1.8

Weak uniqueness for (3.1.19): The uniqueness proof in section 4 of [32], although stated there only for diffusion matrices of a special form, carries over to our situation. For this, the main fact one has to check is the following.

Lemma 3.5.1 *For $\alpha = 1, \ldots, p$, let*

$$\begin{aligned} F_\alpha &:= \{(x^1, \ldots, x^{p-1}) \in S^{p-1} : x^\alpha = 0\} && (\alpha = 1, \ldots, p-1) \\ F_\alpha &:= \{(x^1, \ldots, x^{p-1}) \in S^{p-1} : \textstyle\sum_\beta x^\beta = 1\} && (\alpha = p) \end{aligned} \tag{3.5.1}$$

be the α-face of the $(p-1)$-dimensional simplex S^{p-1}. Then for any solution X to (3.1.19) with $X(0) \in F_\alpha$

$$P[X(t) \in F_\alpha \;\; \forall t \geq 0] = 1. \tag{3.5.2}$$

Proof of Lemma 3.5.1: immediate by the martingale property of solutions to (3.1.19). ∎

In order to show weak uniqueness for (3.1.19) we prove strong uniqueness for the special case where σ is the unique positive symmetric root of w (recall (3.1.7)). By (3.1.55), this σ is Lipschitz continuous on the interior of S^{p-1} and therefore a standard argument gives uniqueness of solutions to (3.1.19) up to the first hitting of a face F_α. By Lemma 3.5.1, the process stays in this face after hitting it. Each face is isomorphic to S^{p-2} and therefore strong uniqueness can be proved by induction. For details we refer to [32].

(a), (b) and (c): By (3.1.55), the effective boundary of K consists of the extremal points of S^{p-1}:

$$\partial_w K = \{e_1, \ldots, e_p\}, \tag{3.5.3}$$

where $e_1 = (1, 0, \ldots, 0)$, $e_{p-1} = (0, \ldots, 0, 1)$ and $e_p = (0, \ldots, 0)$. It follows that for any $f \in \mathcal{C}(K)$

$$(S_\infty f)(x) = \sum_{\alpha=1}^{p} P[X^x(\infty) = e_p] f(e_p), \tag{3.5.4}$$

where the probabilities $P[X^x(\infty) = e_p]$ follow from the martingale property of solutions to (3.1.19). The rest of the assertions is now trivial.

3.5.2 Proof of Example 3.1.9

Uniqueness of solutions to (3.1.19) is proved in the same way as in Example 3.1.8, where this time one needs to check that any solution X starting in $x \in \partial K$ is constant with probability one. By the convexity of K we can without lack of generality assume that $x = 0$ and $y^1 \geq 0$ for all $y \in K$. Then by the martingale property of solutions to (3.1.19) we have

$$P[X^1(t) = 0 \ \ \forall t \geq 0] = 1 \tag{3.5.5}$$

and this implies that almost surely $X(t) \in \partial K$ for all $t \geq 0$ and hence

$$E|X(t)|^2 = \int_0^t E[2g(X(s))]ds = 0 \quad \forall t \geq 0. \tag{3.5.6}$$

Thus we see that almost surely $X(t) = x$ for all $t \geq 0$.

To see that $S_\infty(\mathcal{C}(K)) \subset \mathcal{C}(K)$ and that the class of harmonic functions is given by formula (3.1.60), we can use [22], Proposition 4.2.7 and Theorems 4.2.12 and 4.2.19, where by (3.1.24) the harmonic functions of the process in (3.1.19) are the same as the harmonic functions for Brownian motion. The same references show that (3.1.62) has a unique solution. It follows from (3.1.60) that $T_{\theta,t}(H) \subset H$ and this implies that the boundary distribution is stable against a linear drift. The other assertions in Example 3.1.9 are now readily checked.

3.6 Finite Λ

We state and proof the following lemma, announced in section 3.1.5.

Lemma 3.6.1 *Let X be a shift-invariant solution to (A.5.1) and assume that Λ is finite. Then there exists a K^Λ-valued random variable $X(\infty)$ such that*

$$\lim_{t\to\infty} \sum_i E|X_i(t) - X_i(\infty)|^2 = 0. \tag{3.6.1}$$

Proof of Lemma 3.6.1: Since

$$d\Big(\sum_i X_i(t)\Big) = \sum_i \sigma(X_i(t))dB_i(t), \tag{3.6.2}$$

it is clear that $(\sum_i X_i(t))_{t\geq 0}$ is a bounded martingale, and hence there exists a random variable Z such that

$$\lim_{t\to\infty} E\Big|Z - \sum_i X_i(t)\Big|^2 = 0. \tag{3.6.3}$$

We are therefore done if we can show that

$$\lim_{t\to\infty} E|X_i(t) - X_j(t)|^2 = 0 \qquad \forall i, j \in \Lambda. \tag{3.6.4}$$

We will establish (3.6.4) by showing that each sequence $t_n \to \infty$ has a subsequence $t_{\tilde{n}(m)}$ such that

$$\lim_{p\to\infty} E|X_i(t_{\tilde{n}(p)}) - X_j(t_{\tilde{n}(p)})|^2 = 0 \qquad \forall i, j \in \Lambda. \tag{3.6.5}$$

So let us fix a sequence $t_n \to \infty$. Lemma 5.1 and Remark 5.2 from chapter 4 of [16] imply the existence of a subsequence $t_{n(m)}$ and a process $\tilde{X}$, such that in the sense of weak convergence on path space

$$(X(t_{n(m)} + t))_{t\geq 0} \Rightarrow (\tilde{X}(t))_{t\geq 0} \qquad \text{as } m \to \infty, \tag{3.6.6}$$

where $\tilde{X}$ has sample paths in $\mathcal{D}_{K^\Lambda}[0, \infty)$ and solves the martingale problem for the operator A. Formula (3.6.3) together with the continuous sample paths of $\tilde{X}$ implies that for all $t \geq 0$

$$\sum_i X_i(t_{n(m)} + t) - \sum_i X_i(t_{n(m)}) \Rightarrow 0 \qquad \text{as } m \to \infty, \tag{3.6.7}$$

and therefore

$$\sum_i X_i(t) - \sum_i X_i(0) = 0 \quad \text{a.s.} \qquad (t \geq 0). \tag{3.6.8}$$

Using the fact that $(\sum_i \tilde{X}_i(t))_{t\geq 0}$ is a martingale and the fact that $\tilde{X}$ solves the martingale problem for A, we see that

$$\begin{aligned} E\Big|\sum_i \tilde{X}_i(t) - \sum_i X_i(0)\Big|^2 &= E\Big|\sum_i \tilde{X}_i(t)\Big|^2 - E\Big|\sum_i X_i(0)\Big|^2 \\ &= \sum_i \int_0^t E[tr(w)(\tilde{X}_i(s))]ds. \end{aligned} \tag{3.6.9}$$

We use this, together with (3.6.8) and the fact that $tr(w) \geq 0$ to conclude that

$$E[tr(w)(\tilde{X}_i(t))] = 0 \qquad (i \in \Lambda,\ t \geq 0). \tag{3.6.10}$$

The fact that $\tilde{X}$ solves the martingale problem for the operator A means that for each $f \in \mathcal{C}^2(K^\Lambda)$

$$f(\tilde{X}(t)) - f(\tilde{X}(0)) - \int_0^t (Af)(\tilde{X}(s))ds \tag{3.6.11}$$

is a martingale. Here

$$\begin{aligned}(Af)(\tilde{X}(s)) = &\sum_{ij}\sum_{\alpha} a(j-i)(\tilde{X}_j^\alpha(s) - \tilde{X}_i^\alpha(s))(\tfrac{\partial}{\partial x_i^\alpha} f)(\tilde{X}(s)) \\ &+ \sum_i \sum_{\alpha\beta} w_{\alpha\beta}(\tilde{X}_i(s))(\tfrac{\partial^2}{\partial x_i^\alpha \partial x_i^\beta} f)(\tilde{X}(s)),\end{aligned} \tag{3.6.12}$$

where the second term is zero by (3.6.10). We therefore see that $\tilde{X}$ also solves the martingale problem for the operator

$$\sum_{ij}\sum_{\alpha} a(j-i)(x_j^\alpha(s) - x_i^\alpha(s))\tfrac{\partial}{\partial x_i^\alpha}, \tag{3.6.13}$$

and hence is equal in distribution to a solution of the system of differential equations

$$d\tilde{X}_i(t) = \sum_j a(j-i)(\tilde{X}_j(t) - \tilde{X}_i(t))dt. \tag{3.6.14}$$

By the irreducibility of a this implies that

$$\lim_{t\to\infty} \tilde{X}_i(t) - \tilde{X}_j(t) = 0 \quad \text{a.s.} \qquad (i, j \in \Lambda). \tag{3.6.15}$$

Thus

$$\lim_{t\to\infty}\lim_{m\to\infty} E|X_i(t_{n(m)} + t) - X_j(t_{n(m)} + t)|^2 = 0 \qquad (i, j \in \Lambda). \tag{3.6.16}$$

By a diagonal argument we see that there exist T_m such that for all $s_m \geq T_m$

$$\lim_{m\to\infty} E|X_i(t_{n(m)} + s_m) - X_j(t_{n(m)} + t_m)|^2 = 0 \qquad (i, j \in \Lambda). \tag{3.6.17}$$

In particular, we can find a further subsequence $t_{\tilde{n}(m)}$ such that $t_{\tilde{n}(m)} - t_{n(m)} \geq T_m$, and therefore

$$\lim_{m\to\infty} E|X_i(t_{\tilde{n}(m)}) - X_j(t_{\tilde{n}(m)})|^2 = 0 \qquad (i, j \in \Lambda). \tag{3.6.18}$$

This shows that (3.6.5) holds. ∎

Chapter 4

Large space-time scale universality of hierarchically interacting diffusions

4.1 Introduction and main result

4.1.1 Definitions

We begin by introducing the following objects:

1. ('state space') $K \subset \mathbb{R}^d$ is a compact convex subset of $\mathbb{R}^d$ ($d \geq 1$) with non-empty interior K°, and $\partial K := K \backslash K^\circ$ is its boundary.

2. ('diffusion function') $g : K \to [0, \infty)$ is a function satisfying

$$\begin{aligned} (i) &\quad g = 0 \text{ on } \partial K \\ (ii) &\quad g > 0 \text{ on } K^\circ \\ (iii) &\quad x \mapsto \sqrt{g(x)} \text{ is Lipschitz continuous on } K. \end{aligned} \tag{4.1.1}$$

3. ('index space') Ω_N ($N \geq 2$) is the N-dimensional hierarchical lattice

$$\Omega_N := \Big\{ (\xi_i)_{i \geq 1} : \xi_i \in \{0, 1, \ldots, N-1\},\ \xi_i \neq 0 \text{ finitely often} \Big\}. \tag{4.1.2}$$

With componentwise addition modulo N, Ω_N is a countable Abelian group. The unit element with respect to the group action is the origin $0 = (0, 0, \ldots)$, and Ω_N is equipped with the norm

$$\|\xi\| := \min\{j \geq 0 : \xi_i = 0 \text{ for all } i > j\}. \tag{4.1.3}$$

We fix a function g as in (4.1.1), a constant $c \in (0, 1)$ and a point $\theta \in K$. For each N we consider a family of stochastic processes

$$X^N = \left(X^N(t)\right)_{t \geq 0} = \left(\left\{X^N)_\xi(t)\right\}_{\xi \in \Omega_N}\right)_{t \geq 0} \tag{4.1.4}$$

solving the system of stochastic differential equations

$$dX_\xi^N(t) = \sum_\eta a_N(\eta - \xi)\left[X_\eta^N(t) - X_\xi^N(t)\right]dt + \sqrt{2g(X_\xi^N(t))}\,dB_\xi^N(t), \tag{4.1.5}$$

with initial condition

$$X_\xi^N(0) = \theta \qquad (\xi \in \Omega_N,\ t \geq 0), \tag{4.1.6}$$

where

$$a_N(\xi) := \sum_{k=\|\xi\|}^{\infty} \frac{1}{N^k}\left(\frac{c}{N}\right)^{k-1} \tag{4.1.7}$$

plays the role of an interaction kernel between pairs of components and the collection $(\{B_\xi^N(t)\}_{\xi \in \Omega_N})_{t \geq 0}$ are independent d-dimensional standard Brownian motions.

Because $|\{\xi \in \Omega_N : \|\xi\| \leq k\}| = N^k$, the stochastic differential equations in (4.1.5) can be rewritten as

$$dX_\xi^N(t) = \sum_{k=1}^{\infty}\left(\frac{c}{N}\right)^{k-1}\left[X_\xi^{N,k}(t) - X_\xi^N(t)\right]dt + \sqrt{2g(X_\xi^N(t))}\,dB_\xi^N(t), \tag{4.1.8}$$

where $X_\xi^{N,k}(t)$ is the 'k-block average' around ξ:

$$X_\xi^{N,k}(t) := \frac{1}{N^k}\sum_{\eta : \|\eta - \xi\| \leq k} X_\eta^N(t) \qquad (k \geq 0). \tag{4.1.9}$$

This explains why we choose to write the interaction kernel in the form (4.1.7).

We equip the space K^{Ω_N} with the product topology. In this topology, K^{Ω_N} is a compact separable metrizable space. We write $\mathcal{C}(K^{\Omega_N})$ for the Banach space of continuous real-valued functions on K^{Ω_N}, equipped with the supremum norm. It is shown in Swart [41] that the system in (4.1.5) has a unique strong solution with continuous sample paths. Moreover, X^N solves the martingale problem for a linear operator A on $\mathcal{C}(K^{\Omega_N})$ given by

$$\begin{aligned}(Af)(x) :=& \Big(\sum_{\xi,\alpha}\sum_\eta a_N(\eta - \xi)[x_\eta^\alpha - x_\xi^\alpha]\tfrac{\partial}{\partial x_\xi^\alpha} + \sum_{\xi,\alpha} g(x_\xi)(\tfrac{\partial}{\partial x_\xi^\alpha})^2\Big)f(x) \\ =& \Big(\sum_{\xi,\alpha}\sum_{k=1}^{\infty}\left(\frac{c}{N}\right)^{k-1}[x_\xi^{k,\alpha} - x_\xi^\alpha]\tfrac{\partial}{\partial x_\xi^\alpha} + \sum_{\xi,\alpha} g(x_\xi)(\tfrac{\partial}{\partial x_\xi^\alpha})^2\Big)f(x).\end{aligned} \tag{4.1.10}$$

Here f is a real function on K^{Ω_N}, a typical element $x \in K^{\Omega_N}$ is denoted by

$$x = \{x_\xi\}_{\xi\in\Omega_N} = \{x_\xi^\alpha\}_{\xi\in\Omega_N}^{\alpha=1,\dots,d}, \tag{4.1.11}$$

and x_ξ^k is the k-block averaged variable around ξ:

$$x_\xi^k = \{x_\xi^{k,\alpha}\}^{\alpha=1,\dots,d} := \frac{1}{N^k} \sum_{\eta:\|\eta-\xi\|\le k} x_\eta. \tag{4.1.12}$$

The operator A in (4.1.10) has domain

$$\mathcal{D}(A) := \mathcal{C}^2_{\text{fin}}(K^{\Omega_N}), \tag{4.1.13}$$

the space of twice continuously differentiable functions depending on finitely many coordinates only. Here, for any set $D \subset \mathbb{R}^n$ we denote by $\mathcal{C}^2(D)$ the space of functions $f : D \to \mathbb{R}$ that can be extended to a function in $\mathcal{C}^2(\mathbb{R}^n)$.

4.1.2 Main scaling theorem

We pick the origin as a typical reference point, and we investigate the behavior of block averages around the origin $X_0^{N,k}(t)$ for large N and k. In order to get a sensible result, we must rescale space and time together. It turns out that the process $\hat{X}^{N,k}$ has non-trivial limiting behavior, where

$$\hat{X}^{N,k}(t) := X_0^{N,k}(\beta_{N,k}t) \qquad (t \ge 0) \tag{4.1.14}$$

with

$$\begin{aligned} \beta_{N,k} &:= \sigma_k N^k \\ \sigma_k &:= \sum_{l=0}^{k-1} \frac{1}{c^l} = \frac{c^{-k}-1}{c^{-1}-1}. \end{aligned} \tag{4.1.15}$$

In order to specify the limiting process, we introduce the following objects. Whenever $g : K \to [0,\infty)$ is a continuous function satisfying $g(x) = 0 \Leftrightarrow x \in \partial K$, $c \in [0,\infty)$ is a constant and $x \in K$, we let $Z_x^{g,c}$ be a weak solution of the stochastic differential equation

$$\begin{aligned} dZ(t) &= c[x - Z(t)]\,dt + \sqrt{2g(Z(t))}\,dB(t) \\ Z(0) &= x. \end{aligned} \tag{4.1.16}$$

Existence of $Z_x^{g,c}$ is guaranteed by Theorem 5.4 and Remark 5.5 in Ethier & Kurtz [16].

By g^* we denote the unique solution (continuous on K and twice continuously differentiable on K°) of the equation

$$\begin{aligned} -\tfrac{1}{2}\Delta g^* &= 1 \qquad \text{on } K^\circ \\ g^* &= 0 \qquad \text{on } \partial K, \end{aligned} \tag{4.1.17}$$

with $\Delta = \sum_\alpha (\frac{\partial}{\partial x^\alpha})^2$ the Laplacian.

By $\mathcal{D}_K[0,\infty)$ we denote the space of cadlag functions from $[0,\infty)$ to K, equipped with the Skorohod topology. We use the symbol $\Rightarrow$ to denote weak convergence of processes in path space, i.e., the weak convergence of their laws as probability measures on $\mathcal{D}_K[0,\infty)$.

$\underline{d=1}$: For one-dimensional K, exact results have been derived about the asymptotic behavior of $\hat{X}^{N,k}$. Let $K=[0,1]$. Let $\mathcal{H}$ be the class of Lipschitz continuous functions $g:[0,1]\to[0,\infty)$ satisfying $g(x)=0 \Leftrightarrow x\in\{0,1\}$. Then (4.1.16) has a unique strong solution for every $g\in\mathcal{H}$, $x\in K$ and $c\in[0,\infty)$. We cite the following result from Dawson and Greven [10].

Proposition 4.1.1 *If $g\in\mathcal{H}$, then for every $k\geq 0$*

$$\hat{X}^{N,k} \Rightarrow Z_\theta^{g_k,c_k} \qquad \text{as } N\to\infty, \tag{4.1.18}$$

where $c_k := \sigma_k c^k$, and $g_k\in\mathcal{H}$ is the function

$$g_k := \sigma_k(F_{c^{k-1}}\circ\cdots F_{c^1}\circ F_{c^0})g. \tag{4.1.19}$$

Here $F_c:\mathcal{H}\mapsto\mathcal{H}$ is a renormalization transformation given by

$$(F_c g)(x) := \int_{[0,1]} g(y)\nu_x^{g,c}(dy) \qquad (x\in[0,1]), \tag{4.1.20}$$

where $\nu_x^{g,c}$ is the unique equilibrium distribution of (4.1.16).

In Baillon *et al.* [1] the behavior of the function g_k was studied in the limit as $k\to\infty$. The main result of that paper is the following:

Proposition 4.1.2 *For any $g\in\mathcal{H}$*

$$\lim_{k\to\infty}\sup_{x\in[0,1]} |g_k(x)-g^*(x)| = 0, \tag{4.1.21}$$

where the function g^* is given by

$$g^*(x) := x(1-x) \qquad (x\in[0,1]). \tag{4.1.22}$$

Since

$$c_k \to c^* = \frac{c}{1-c} \qquad \text{as } k \to \infty, \tag{4.1.23}$$

we may combine Propositions 4.1.1 and 4.1.2 to obtain (see [21])

$$\hat{X}^{N,k}(t) \Rightarrow Z_\theta^{g^*,c^*} \qquad N \to \infty,\ k \to \infty, \tag{4.1.24}$$

where the limits need to be taken in the order indicated. The fact that large blocks are always governed by the same diffusion function g^*, regardless of the diffusion function g for single components, is described by saying that systems of the form (4.1.5) exhibit *universal behavior on large space-time scales*.

$\underline{d \geq 2}$: In den Hollander and Swart [21] we conjectured that the result in (4.1.24) generalizes to higher-dimensional K, where the function g^* is given by (4.1.17). However, we were only able to prove some partial results in this direction. The main difficulty we encountered was that it is very hard to prove for a function g that the transformations F_{c^k} are well-defined on g *and* on all its iterates $F_0 g$, $(F_c \circ F_0)g$ and so on. This requires that uniqueness holds for solutions of (4.1.16) for all $x \in K$, and not only for g itself, but also for its iterates. This is often already a problem for $g = g^*$

In the present paper, we do not attempt to prove seperate limit theorems for $N \to \infty$ and $k \to \infty$ such as Proposition 4.1.1 and Proposition 4.1.2, but instead let N and k tend to infinity together. In the case of one-dimensional K, the Propositions 4.1.1 and 4.1.2 already imply, through formula (4.1.24), that it is possible to choose N_i, k_i, tending to infinity as $i \to \infty$, such that

$$\hat{X}^{N_i,k_i}(t) \Rightarrow Z_\theta^{g^*,c^*} \qquad \text{as } i \to \infty. \tag{4.1.25}$$

In this paper, we try to generalize this result to higher-dimensional K. Moreover, we investigate how N_i and k_i must be chosen for the convergence in (4.1.24) to hold.

In what follows, we fix integers $N_i \geq 2$, $k_i \geq 1$ $(i \in \mathbb{N})$ tending to infinity as $i \to \infty$. We write

$$\hat{X}^i := \hat{X}^{N_i,k_i} \qquad \beta_i := \beta_{N_i,k_i} \qquad \hat{\mathcal{F}}_t^i := \mathcal{F}_{\beta_i t}^i, \tag{4.1.26}$$

where $(\mathcal{F}_t^i)_{t\geq 0}$ is the filtration generated by the process X^{N_i,k_i}. Unfortunately, we do not know yet how to prove (4.1.25) in the case of higher-dimensional K. However, we can show that the drift and the diffusion rate of the process $\hat{X}^{N_i,k_i}$ converge to the functions $x \mapsto c^*(\theta - x)$ and $x \mapsto g^*(x)$, respectively. The following scaling theorem is our main result.

Theorem 4.1.3 *Let $N_i \geq 2$, $k_i \geq 1$ ($i \in \mathbb{N}$) be integers tending to infinity as $i \to \infty$ such that*

$$\lim_{i\to\infty} \frac{k_i}{\log N_i} = 0, \tag{4.1.27}$$

and assume that $g^* \in \mathcal{C}^1(K)$. Then there exist $(\hat{\mathcal{F}}^i_t)_{t\geq 0}$-adapted processes $\hat{B}^i = (\hat{B}^i(t))_{t\geq 0}$ and $\hat{G}^i = (\hat{G}^i(t))_{t\geq 0}$ with sample paths in $\mathcal{D}_{\mathbb{R}^d}[0,\infty)$, $\mathcal{D}_{\mathbb{R}}[0,\infty)$, respectively, such that for each $f \in \mathcal{C}^2(K)$ the process $(M^i(t))_{t\geq 0}$, given by

$$\begin{aligned} M^i(t) := f(\hat{X}^i(t)) - \int_0^t \Big\{ & \sum_\alpha \hat{B}^{i,\alpha}(s)\Big(\tfrac{\partial}{\partial x^\alpha} f\Big)\Big(\hat{X}^i(s)\Big) \\ & + \hat{G}^i(s)\Big(\sum_\alpha (\tfrac{\partial}{\partial x^\alpha})^2 f\Big)\Big(\hat{X}^i(s)\Big)\Big\}\, ds \end{aligned} \tag{4.1.28}$$

is an $(\hat{\mathcal{F}}^i_t)_{t\geq 0}$-martingale. Moreover, for each $T > 0$

$$\lim_{i\to\infty} E\left[\int_0^T \left|\hat{B}^i(t) - c^*(\theta - \hat{X}^i(t))\right|^2 dt\right] = 0. \tag{4.1.29}$$

and

$$\lim_{i\to\infty} E\left[\left|\int_0^T \Big(\hat{G}^i(t) - g^*(\hat{X}^i(t))\Big)\, dt\right|^2\right] = 0. \tag{4.1.30}$$

Formula (4.1.28) identifies $\hat{B}^i$ and $\hat{G}^i$ as the drift and the diffusion rate of the process $\hat{X}^i$. Thus, formulas (4.1.29) and (4.1.30) show that these local characteristics of the process $\hat{X}^i$ converge, as $i \to \infty$, and that their limits are *universal* in the diffusion function g for single components. The fact that this happens for all N_i, k_i satisfying condition (4.1.27) is new even in the case of one-dimensional K. Theorem 4.1.3 is a universality result of the type we were originally after in den Hollander and Swart [21].

The author believes that also the convergence in (4.1.25) holds under condition (4.1.27), but there are at present two technical obstacles to proving a result of that form.

The first difficulty comes from the fact that uniqueness of solutions to (4.1.16) for arbitrary $g = g^*$, $c = c^*$ and $x = \theta$ remains an open problem. The following partial results are known.

1. Strong uniqueness is known to hold for $\theta \in K^\circ$, c^* sufficiently large, and K satisfying mild regularity conditions (Theorem 1.9 in Den Hollander & Swart [21]).

2. Weak uniqueness is known to hold for $K = \{x \in \mathbb{R}^d : |x|^2 \leq 1\}$, in which case $g^*(x) = (1 - |x|^2)/d$, and θ and c^* arbitrary (Theorem 1.10 in Den Hollander & Swart [21]).

The second difficulty comes from the fact that even if uniqueness of $Z_\theta^{g^*,c^*}$ is known, the results in Theorem 4.1.3 are not sufficient to conclude that $\hat{X}^i$ converges to $Z_\theta^{g^*,c^*}$. In particular, the type of convergence in (4.1.30) is not sufficient to show tightness of $(\hat{X}^i)_{i\in\mathbb{N}}$ in the topology of weak convergence in path space $\mathcal{D}_K[0,\infty)$. If instead of (4.1.30) we would have

$$\lim_{i\to\infty} E\Big[\int_0^T \Big|\hat{G}^i(t) - g^*(\hat{X}^i(t))\Big|^2 dt\Big] = 0, \tag{4.1.31}$$

then tightness of $(\hat{X}^i)_{i\in\mathbb{N}}$ would follow by Theorem 9.4 from Chapter 3 in Ethier & Kurtz [16]. The author believes (4.1.31) to hold, but, as explained in Section 4.5, the techniques in the present paper are not sufficient to show this. In future work we hope to establish the convergence of $(\hat{X}^i)_{i\in\mathbb{N}}$, either by showing (4.1.31) to prove weak convergence in path space, or by using (4.1.30) to prove convergence in some weaker sense.

4.2 Identification of the drift and diffusion rate of $\hat{X}^i$

In this section, we identify processes $\hat{B}^i$ and $\hat{G}^i$ such that (4.1.28) holds. In the following sections we then prove (4.1.29) and (4.1.30), which completes the proof of Theorem 4.1.3.

Lemma 4.2.1 *For each $f \in \mathcal{C}^2(K)$ the process $(M^i(t))_{t\geq 0}$, given by*

$$\begin{aligned} M^i(t) := f(\hat{X}^i(t)) - \int_0^t \Big\{ &\sum_\alpha \hat{B}^{i,\alpha}(s)\Big(\tfrac{\partial}{\partial x^\alpha} f\Big)\Big(\hat{X}^i(s)\Big) \\ &+ \hat{G}^i(s)\Big(\sum_\alpha (\tfrac{\partial}{\partial x^\alpha})^2 f\Big)\Big(\hat{X}^i(s)\Big)\Big\}\, ds \end{aligned} \tag{4.2.1}$$

with

$$\begin{aligned} \hat{B}^i(t) &:= \sigma_{k_i} c^{k_i} \sum_{n=0}^\infty \Big(\frac{c}{N_i}\Big)^n \Big[X_0^{N_i,k_i+n+1}(\beta_i t) - X_0^{N_i,k_i}(\beta_i t)\Big] \\ \hat{G}^i(t) &:= \sigma_{k_i} \frac{1}{N_i^{k_i}} \sum_{\xi:\|\xi\|\leq k_i} g(X_\xi^{N_i}(\beta_i t)). \end{aligned} \tag{4.2.2}$$

is an $(\hat{\mathcal{F}}_t^i)_{t\geq 0}$-martingale.

Proof of Lemma 4.2.1: Fix a function $f \in \mathcal{C}^2(K)$ and for $N \geq 2$ and $k \geq 1$ define $f^{N,k} \in \mathcal{C}^2_{\text{fin}}(K^\Lambda)$ by

$$f^{N,k}(x) := f(x_0^k) = f\Big(\frac{1}{N^k}\sum_{\xi:\|\xi\|\leq k} x_\xi\Big). \tag{4.2.3}$$

Then the process $(M^{N,k}(t))_{t\geq 0}$ is a martingale, where

$$\begin{aligned} M^{N,k}(t) &:= f(X_0^{N,k}(\beta_{N,k}t)) - \int_0^{\beta_{N,k}t} (Af^{N,k})(X^N(s))\,ds \\ &= f(\hat{X}^{N,k}(t)) - \int_0^t \beta_{N,k}(Af^{N,k})(X^N(s))\,ds. \end{aligned} \tag{4.2.4}$$

Here

$$\begin{aligned} \beta_{N,k}(Af^{N,k})(x) = \sigma_k N^k \sum_{\eta,\alpha}\sum_{l=1}^{\infty}\Big(\frac{c}{N}\Big)^{l-1}[x_\eta^{l,\alpha} - x_\eta^\alpha]\tfrac{\partial}{\partial x_\eta^\alpha} f(x_0^k) \\ +\sigma_k N^k \sum_{\eta,\alpha} g(x_\eta)(\tfrac{\partial}{\partial x_\eta^\alpha})^2 f(x_0^k). \end{aligned} \tag{4.2.5}$$

The first term on the right-hand side in (4.2.5) can be written as

$$\begin{aligned} &\sigma_k N^k \sum_{\alpha}\sum_{l=1}^{\infty}\Big(\frac{c}{N}\Big)^{l-1}\frac{1}{N^k}\sum_{\eta:\|\eta\|\leq k}[x_\eta^{l,\alpha} - x_\eta^\alpha](\tfrac{\partial}{\partial x^\alpha} f)(x_0^k) \\ &= \sigma_k N^k \sum_{\alpha}\sum_{l=k+1}^{\infty}\Big(\frac{c}{N}\Big)^{l-1}[x_0^{l,\alpha} - x_0^{k,\alpha}](\tfrac{\partial}{\partial x^\alpha} f)(x_0^k) \\ &= \sigma_k c^k \sum_{\alpha}\sum_{n=0}^{\infty}\Big(\frac{c}{N}\Big)^{n}[x_0^{k+n+1,\alpha} - x_0^{k,\alpha}](\tfrac{\partial}{\partial x^\alpha} f)(x_0^k), \end{aligned} \tag{4.2.6}$$

where in the first equality we use that for $l \leq k$

$$\begin{aligned} \frac{1}{N^k}\sum_{\eta:\|\eta\|\leq k}[x_\eta^l - x_\eta] &= \frac{1}{N^k}\sum_{\eta:\|\eta\|\leq k}\frac{1}{N^l}\sum_{\zeta:\|\zeta-\eta\|\leq l}[x_\zeta - x_\eta] \\ &= \frac{1}{N^{k+l}}\sum_{\substack{\eta:\|\eta\|\leq k\\ \zeta:\|\zeta\|\leq k\\ \|\zeta-\eta\|\leq l}}[x_\zeta - x_\eta] = 0 \end{aligned} \tag{4.2.7}$$

and for $l > k$

$$\frac{1}{N^k}\sum_{\eta:\|\eta\|\leq k}[x_\eta^l - x_\eta] = \frac{1}{N^k}\sum_{\eta:\|\eta\|\leq k}[x_0^l - x_\eta] = [x_0^l - x_0^k]. \tag{4.2.8}$$

The second term on the right-hand side in (4.2.5) can be written as

$$\sigma_k N^k \frac{1}{N^{2k}} \sum_{\eta:\|\eta\|\le k} g(x_\eta)(\sum_\alpha (\tfrac{\partial}{\partial x^\alpha})^2 f)(x_0^k). \tag{4.2.9}$$

Inserting (4.2.9) and (4.2.6) into (4.2.5), inserting (4.2.5) in (4.2.4) and defining $M^i := M^{N_i,k_i}$ we arrive at Lemma 4.2.1. ∎

4.3 Convergence of the drift

We start by giving an upper bound on the speed with which block averages change in time. In the following lemma we consider the process

$$\left(X^N(t+s)\right)_{s\ge 0} = \left(\left\{X_\xi^N(t+s)\right\}_{\xi\in\Omega_N}\right)_{s\ge 0}, \tag{4.3.1}$$

conditioned on the event

$$X^{N_i}(\beta_i t) = x, \tag{4.3.2}$$

with $x \in K^{\Omega_N}$. We choose a regular version of the conditional expectation

$$E_x[\,\cdot\,] := E[\,\cdot\,|X^N(t) = x], \tag{4.3.3}$$

with the property that under the conditional law, the process in (4.3.1) solves the martingale problem for the operator A in (4.1.10) with initial condition x.

Lemma 4.3.1 *There exists a constant M such that for all integers $N \ge 2$ and $k \ge 1$, for all $t, s \ge 0$ and for all $x \in K^{\Omega_N}$*

$$E\left[\left|X_0^{N,k}(t+s) - x_0^k\right|^2 \,\middle|\, X^N(t) = x\right] \le M\frac{s}{N^k}. \tag{4.3.4}$$

Proof of Lemma 4.3.1: A calculation similar to the proof of Lemma 4.2.1 then gives

$$\begin{aligned}
E_x\left[\left|X_0^{N,k}(t+s) - x_0^k\right|^2\right] &= \int_0^s du\, \frac{1}{N^{2k}} \sum_{\xi:\|\xi\|\le k} 2dE[g(X_\xi^N(u))] \\
&+ \int_0^s du \sum_{l=k+1}^\infty \left(\frac{c}{N}\right)^{l-1} \\
&\qquad 2E_x\left[\sum_\alpha \left[X_0^{N,l,\alpha}(t+u) - X_0^{N,k,\alpha}(t+u)\right]\left[X_0^{N,k,\alpha}(t) - x_0^{k,\alpha}\right]\right] \\
&\le \frac{2ds}{N^k}\|g\|_\infty + 2R^2 s\left(\frac{c}{N}\right)^k \sum_{n=0}^\infty \left(\frac{c}{N}\right)^n \\
&\le \left(2d\|g\|_\infty + 4R^2\right)\frac{s}{N^k},
\end{aligned} \tag{4.3.5}$$

where R is a constant such that $|x-y| \leq R$ for all $x, y \in K$ and in the last step we use that $N \geq 2$ and $c < 1$. ∎

Proof of formula (4.1.29): Since the terms with $n \geq 1$ in (4.2.2) tend to zero uniformly as $i \to \infty$, it suffices to show that

$$\lim_{i\to\infty} E\int_0^T \left|\sigma_{k_i} c^{k_i}\left[X_0^{N_i,k_i+1}(\beta_i t) - \hat{X}^i(t)\right] - c^*[\theta - \hat{X}^i(t)]\right|^2 dt = 0. \tag{4.3.6}$$

Since $\sigma_k c^k \to c^*$as $k \to \infty$ (recall (4.1.15) and 4.1.23)), it thus suffices to show that

$$\lim_{i\to\infty} E\int_0^T \left|X_0^{N_i,k_i+1}(\beta_i t) - \theta\right|^2 dt = 0. \tag{4.3.7}$$

We use Lemma 4.3.1 to estimate ($X^{N_i,k_i+1}(0) = \theta$)

$$E\int_0^T \left|X_0^{N_i,k_i+1}(\beta_i t) - \theta\right|^2 dt \leq T \sup_{0\leq s\leq \beta_i T} E\left[\left|X_0^{N_i,k_i+1}(s) - \theta\right|^2\right] \leq M\frac{\beta_i T^2}{N_i^{k_i+1}}. \tag{4.3.8}$$

Since T is fixed, the right-hand side tends to zero provided that

$$\lim_{i\to\infty} \frac{\beta_i}{N_i^{k_i+1}} = 0. \tag{4.3.9}$$

Inserting $\beta_i = \sigma_{k_i} N_i^{k_i}$ and $\sigma_{k_i} \sim c^{-k_i}(c/(1-c))$, we find that this condition amounts to

$$\lim_{i\to\infty} c^{k_i} N_i = \infty. \tag{4.3.10}$$

But the latter holds for any $c \in (0, 1)$ because of condition (4.1.27). ∎

4.4 Convergence of the diffusion rate

4.4.1 Strategy of the proof

In this section the essential ideas behind Theorem 4.1.3 will have to come in. In particular, we will need to explain how the universal large space-time diffusion function g^* arises and why the scaling of time with the factor $\sigma_{k_i} N^{k_i}$ is the correct one. Before we embark on the calculations that will give us the convergence in (4.1.30), we outline the heuristics of the proof.

STEP 1: We fix a $t \geq 0$ and look at the process

$$\left(X_\xi^{N_i}(\beta_i t + s)\right)_{\xi:\|\xi\|\leq k_i - 1,\ s\in[0,T_i]}, \tag{4.4.1}$$

with

$$\sigma_{k_i-1} N_i^{k_i-1} \ll T_i \ll N_i^{k_i}. \tag{4.4.2}$$

Thus, we consider the evolution of a $(k_i - 1)$-block on a time scale that is long with respect to $\sigma_{k_i-1} N_i^{k_i-1}$ (the presumed time scale of the $(k_i - 1)$-block average), but short with respect to $N_i^{k_i}$. Note that condition (4.4.2) can be met because of condition (4.1.27).

The assumption that $T_i \ll N_i^{k_i}$ allows us to simplify the stochastic differential equations in (4.1.8). First, we can neglect the terms in the summation with $k \geq k_i + 1$, because they are of order $N_i^{-k_i}$ and will not be felt on times $T_i \ll N_i^{k_i}$. Second, according to Lemma 4.3.1, the block average $X_0^{N_i,k_i}$ can be considered as essentially fixed over times $T_i \ll N_i^{k_i}$, and hence we expect that the time evolution of the system in (4.4.1) can be approximated by the equations

$$\begin{aligned} dX_\xi^{N_i}(\beta_i t + s) = & \left(\frac{c}{N_i}\right)^{k_i-1} \left[X_\xi^{N_i,k_i}(\beta_i t) - X_\xi^{N_i}(\beta_i t + s)\right] ds \\ & + \sum_{k=1}^{k_i-1} \left(\frac{c}{N_i}\right)^{k-1} \left[X_\xi^{N_i,k}(\beta_i t + s) - X_\xi^{N_i}(\beta_i t + s)\right] ds \\ & + \sqrt{2g(X_\xi^{N_i}(\beta_i t + s))} dB_\xi^i(\beta_i t + s) \qquad (s \geq 0,\ \|\xi\| \leq k_i - 1). \end{aligned} \tag{4.4.3}$$

Next comes the essential point in the argument. We expect that the condition $\sigma_{k_i-1} N_i^{k_i-1} \ll T_i$ is sufficient to guarantee that solutions of (4.4.3) reach *equilibrium* on the time scale T_i, conditional to the k_i-block average $X_\xi^{N_i,k_i}(\beta_i t)$. The system in (4.1.5) as a whole does not have a true equilibrium distribution; instead, it was shown in Swart [41] that the distribution of the system tends to a mixture of trivial extremal measures as $t \to \infty$. However, as was recognized by Dawson and Greven in [10], the system in (4.1.5) goes through a series of 'local equilibria' as time tends to infinity, where k_i-blocks of ever larger size reach a (temporary and approximate) 'local' equilibrium at times of the appropriate order of magnitude. It is from the properties of these local equilibria that our result will follow.

STEP 2: Let us condition the system on

$$X_0^{N_i,k_i}(\beta_i t) = \hat{\theta}, \tag{4.4.4}$$

and assume that the system in (4.4.1), conditioned on (4.4.4), is in equilibrium. For $\|\xi\| \leq k_i - 1$ and $\|\eta\| \leq k_i$, we define the covariance function

$$C_s(\xi - \eta) := \mathrm{Cov}\left(X_\xi^{N_i}(\beta_i t + s), X_\eta^{N_i}(\beta_i t + s)\right), \tag{4.4.5}$$

where the covariance of two K-valued random variables X and Y is defined as

$$\mathrm{Cov}(X,Y) := E[X \cdot Y] - E[X] \cdot E[Y] \tag{4.4.6}$$

with $x \cdot y := \sum_\alpha x^\alpha y^\alpha$ the usual inner product on $\mathbb{R}^d$. A covariance calculation as in Swart [41] gives that for $\|\xi\| \le k_i$

$$\begin{aligned}\tfrac{\partial}{\partial s} C_s(\xi) &= \sum_\eta a_{N_i}^{k_i-1}(\eta-\xi)[C_s(\eta) - C_s(\xi)] \\ &+ 2d\delta_{0,\xi} E[g(X_0^{N_i}(\beta_i t + s))] - 2\Big(\frac{c}{N_i}\Big)^{k_i} C_s(\xi),\end{aligned} \tag{4.4.7}$$

where a_N^k is the k-block interaction kernel

$$a_N^k(\xi) := \sum_{l=\|\xi\|}^{k} \frac{1}{N^l}\Big(\frac{c}{N}\Big)^{l-1}. \tag{4.4.8}$$

Using our assumption about local equilibrium, we set $\frac{\partial}{\partial s} C_s(\xi) = 0$ in (4.4.7) and we assume that $E[g(X_\xi^{N_i}(\beta_i t + s))]$ does not depend on s. Now we can solve $C_s(\xi)$ in terms of $E[g(X_\xi^{N_i}(\beta_i t + s))]$ and a random walk on

$$\Delta_i := \{\xi \in \Omega_{N_i} : \|\xi\| \le k_i - 1\} \tag{4.4.9}$$

that jumps from site ξ to site η with rate $a_{N_i}^{k_i-1}(\eta-\xi)$ and that is killed in each site with rate $(\frac{c}{N_i})^{k_i}$. Indeed, denoting by $P_t^i(\eta-\xi)$ the probability that this random walk moves from site ξ to site η in time t, we have the representation

$$C_s(\xi) = d\, E[g(X_0^{N_i}(\beta_i t + s))] \int_0^\infty P_t^i(\xi) dt. \tag{4.4.10}$$

(Note that with probability one the random walk is eventually killed, so that the integral on the right-hand side is finite.) Picking $\xi = 0$, we get

$$\mathrm{Var}(X_0^{N_i}(\beta_i t + s)) = d\mu_i E[g(X_0^{N_i}(\beta_i t + s))] \tag{4.4.11}$$

with

$$\mu_i := \int_0^\infty P_t^i(0) dt \tag{4.4.12}$$

the expected time the random walk starting in 0 spends at the origin.

STEP 3: It turns out that we can also express the expectation of any harmonic function of $X_\xi^{N_i}(\beta_i t + s)$ in terms of the above random walk. Indeed, we have the representation (see Swart [41], Lemma 3.1.6 in this dissertation)

$$E[f(X_0^{N_i}(\beta_i t + s))] = E\Big[f\Big(\hat\theta + \sum_\xi P_s^i(\xi)[X_\xi^{N_i}(\beta_i t) - \hat\theta]\Big)\Big] \tag{4.4.13}$$

for any function $f \in \mathcal{C}^2(K^\circ) \cap \mathcal{C}(K)$ satisfying

$$\sum_\alpha (\tfrac{\partial}{\partial x^\alpha})^2 f(x) = 0 \qquad (x \in K^\circ). \tag{4.4.14}$$

Formula (4.4.13) says that harmonic functions of a component evolve under the semigroup associated with the evolution in (4.4.3) as if the diffusion function g is zero.

The assumption of local equilibrium now leads to the relation

$$E[f(X_0^{N_i}(\beta_i t + s))] = f(\hat{\theta}), \tag{4.4.15}$$

which may be described by saying that the 'harmonic mean' of $X_0^{N_i}(\beta_i t + s)$ is $\hat{\theta}$. We next note that the function

$$x \mapsto dg^*(x) + |x - \hat{\theta}|^2 \tag{4.4.16}$$

is harmonic. Therefore, combining (4.4.15) and (4.4.11), we find that

$$\mu_i E[g(X_0^{N_i}(\beta_i t + s))] = g^*(\hat{\theta}) - E[g^*(X_0^{N_i}(\beta_i t + s))]. \tag{4.4.17}$$

<u>STEP 4</u>: We will show that $\mu_i \sim \sigma_{k_i}$ $(i \to \infty)$. Hence (4.4.17) becomes

$$\sigma_{k_i} E[g(X_0^{N_i}(\beta_i t + s))] \sim g^*(\hat{\theta}) - E[g^*(X_0^{N_i}(\beta_i t + s))] \qquad (i \to \infty). \tag{4.4.18}$$

Since σ_{k_i} tends to infinity and the right-hand side of (4.4.18) is bounded by $\|g^*\|_\infty$, it follows that $E[g(X_0^{N_i}(\beta_i t + s))]$ tends to zero as $i \to \infty$. This means that, with high probability, the components $X_\xi^{N_i}(\beta_i t + s)$ of the system are concentrated near the boundary of K, i.e., the system clusters. Since g^* is continuous on K and zero on ∂K, it follows that also $E[g^*(X_0^{N_i}(\beta_i t + s))]$ tends to zero as $i \to \infty$. Hence, using (4.4.18) once more we see that

$$\lim_{i \to \infty} \sigma_{k_i} E[g(X_0^{N_i}(\beta_i t + s))] = g^*(\hat{\theta}). \tag{4.4.19}$$

<u>STEP 5</u>: We now consider the k_i-block $\{\xi \in \Omega_{N_i} : \|\xi\| \le k_i\}$. The $(k_i - 1)$-blocks that the k_i-block consists of, N_i in total, all reach equilibrium on the time scale T_i, and they do so independently of each other. Hence we expect a law of large numbers to apply. In particular, we expect that

$$\lim_{i \to \infty} \mathrm{Var}\Big(N_i^{-k_i} \sum_{\xi : \|\xi\| \le k_i} \sigma_{k_i} g(X_\xi^{N_i}(\beta_i t + s))\Big) = 0. \tag{4.4.20}$$

Inspecting the definition of $\hat{G}^i$ in (4.2.2), we see that (4.4.19) and (4.4.20) imply the convergence of $\hat{G}^i$ to $g^*(\hat{X}^t)$, as claimed in (4.1.30).

In what follows we will have to turn the heuristic reasoning in Steps 1–5 into a solid proof. The main difficulty we have to overcome is that, as i tends to infinity, not only does T_i (our time scale) tend to infinity, so do N_i and k_i. We therefore cannot really say that the system

$$\Big(X_\xi^{N_i}(\beta_i t + s)\Big)_{\xi:\|\xi\|\leq k_i-1,\ s\in[0,T_i]} \tag{4.4.21}$$

tends to equilibrium as $T_i \to \infty$, because the space it lives on changes with i. We will have to find a way to measure how close the system is to equilibrium. We will do so by looking at the system at an exponentially distributed random time with mean T_i, rather than at a fixed time T_i, i.e., we effectively take a Laplace transform with respect to the time variable. For the distribution of the system at this random time we will derive an equilibrium equation with an error term. Extending a technique first used in Den Hollander & Swart [21], we will reformulate the heuristic line of reasoning that led us to formula (4.4.19) in such a way that it depends on the equilibrium equation only. In this way we are able to control the errors that were made to derive (4.4.19). In order to justify also (4.4.20) in some rigorous form, we condition the system on one $(k_i - 1)$-block and show that this has a negligable effect on the behavior of other $(k_i - 1)$-blocks. In this way we are finally able to justify formula (4.1.30) rigorously.

4.4.2 Definitions

For each $N \geq 0$ and $x \in K^{\Omega_N}$ let X^N be a solution of the system of stochastic differential equations in (4.1.5) with initial condition

$$X_\xi^N(0) = x_\xi \qquad (\xi \in \Omega_N), \tag{4.4.22}$$

and for each $i \in \mathbb{N}$ pick $\xi_i \in \Omega_{N_i}$ such that

$$\|\xi_i\| = k_i, \tag{4.4.23}$$

and write $E_\omega[\,\cdot\,]$ for the conditional expectation

$$E_\omega[\,\cdot\,] := E\Big[\,\cdot\,\Big|\Big(X_{\xi_i}^{N_i}(t)\Big)_{t\geq 0} = (\omega(t))_{t\geq 0}\Big], \tag{4.4.24}$$

where $\omega \in \mathcal{C}_{x_{\xi_i},K}[0,\infty)$, the space of all continuous functions $\omega : [0,\infty) \to K$ satisfying $\omega(0) = x_{\xi_i}$. We choose a regular version of $E_\omega[\,\cdot\,]$ with the property that for every $\omega \in \mathcal{C}_{x_{\xi_i},K}[0,\infty)$ and for every $f \in \mathcal{C}^2_{\text{fin}}(K^{\Omega_{N_i}\setminus\{\xi_i\}})$ and under the conditional law, the process $(M(t))_{t\geq 0}$ is a martingale, where

$$\begin{aligned} M(t) &:= f(X^{N_i}(t)) - f(X^{N_i}(0)) \\ &- \int_0^t \Big\{ \sum_{\xi\neq\xi_i,\alpha} \sum_{k=1}^{\infty} \Big(\frac{c}{N_i}\Big)^{k-1} \Big[X_\xi^{N_i,k,\alpha}(s) - X_\xi^{N_i,\alpha}(s)\Big]\big(\tfrac{\partial}{\partial x_\xi^\alpha} f\big)(X^{N_i}(s))\Big\} ds \\ &- \int_0^t \Big\{ \sum_{\xi\neq\xi_i} g(X_\xi^{N_i}(s))\Big(\sum_\alpha \big(\tfrac{\partial}{\partial x_\xi^\alpha}\big)^2 f\Big)(X^{N_i}(s))\Big\} ds. \end{aligned} \tag{4.4.25}$$

Here $X_\xi^{N_i,k}(t)$ as usual denotes the block average

$$X_\xi^{N_i,k}(t) := \frac{1}{N_i^k} \sum_{\eta:\|\eta-\xi\|\leq k} X_\eta^{N_i}(t) \qquad (k \geq 0), \tag{4.4.26}$$

where in the sum over η the term with $\eta = \xi_i$ is in this case given by

$$X_{\xi_i}^{N_i}(t) = \omega(t) \qquad (t \geq 0). \tag{4.4.27}$$

For each $i \in \mathbb{N}$ we introduce a stopping time τ_i (possibly defined on an extension of our probability space) independent of the process X^{N_i} and of the Brownian motions $(\{B_\xi^{N_i}(t)\}_{\xi\in\Omega_N})_{t\geq 0}$ and exponentially distributed with mean λ_i. Here the λ_i are positive numbers that we will later choose in a suitable way.

In what follows, R is a constant such that $|x-y| \leq R$ for all $x, y \in K$.

4.4.3 Block immobility

Lemma 4.4.1 *For $i \in \mathbb{N}$ let X^{N_i} be a solution of (4.1.5) with initial condition (4.4.22). There exists a constant M such that for all $i \in \mathbb{N}$, $k \geq 1$, $x \in K^{\Omega_{N_i}}$, $\omega \in \mathcal{C}_{\xi_i,K}[0,\infty)$ and $t \geq 0$*

$$E_\omega\Big[\Big|X_0^{N_i,k}(t) - x_0^k\Big|^2\Big] \leq M\frac{t}{N_i^k}. \tag{4.4.28}$$

Proof of Lemma 4.4.1: We first treat the case that $k \geq k_i$. In this case

$$X_0^{N_i,k}(t) = \frac{1}{N_i^k}\Big(\omega(t) + \sum_{\substack{\xi:\|\xi\|\leq k\\ \xi\neq\xi_i}} X_\xi^{N_i}(t)\Big), \tag{4.4.29}$$

and

$$\begin{aligned} \Big|X_0^{N_i,k}(t) - x_0^k\Big|^2 &= \frac{1}{N_i^{2k}}\Big|\sum_{\xi:\|\xi\|\leq k} (X_\xi^{N_i}(t) - x_\xi)\Big|^2 \\ &\leq \frac{1}{N_i^{2k}}\Big|\sum_{\substack{\xi:\|\xi\|\leq k\\ \xi\neq\xi_i}} (X_\xi^{N_i}(t) - x_\xi)\Big|^2 + \Big(\frac{2N_i^k - 1}{N_i^{2k}}\Big)R^2. \end{aligned} \tag{4.4.30}$$

Let $f \in \mathcal{C}^2_{\text{fin}}(K^{\Omega_{N_i} \setminus \{\xi_i\}})$ be given by

$$f(y) := \frac{1}{N_i^{2k}} \Big| \sum_{\substack{\xi : \|\xi\| \le k \\ \xi \neq \xi_i}} (y_\xi - x_\xi) \Big|^2. \tag{4.4.31}$$

Then the fact that the process M in (4.4.25) is a martingale implies that

$$\begin{aligned}
&E_\omega \Bigg[\frac{1}{N_i^{2k}} \Big| \sum_{\substack{\xi : \|\xi\| \le k \\ \xi \neq \xi_i}} (X_\xi^{N_i}(t) - x_\xi) \Big|^2 \Bigg] \\
&= \frac{1}{N_i^{2k}} E_\omega \int_0^t \Big\{ \sum_{\xi \neq \xi_i, \alpha} \sum_{l=1}^\infty \Big(\frac{c}{N_i} \Big)^{l-1} \Big[X_\xi^{N_i, l, \alpha}(s) - X_\xi^{N_i, \alpha}(s) \Big] \\
&\qquad\qquad 1_{\{\|\xi\| \le k\}}\, 2 \sum_{\substack{\eta : \|\eta\| \le k \\ \eta \neq \eta_i}} \Big(X_\eta^{N_i, \alpha}(s) - x_\xi \Big) \Big\} ds \\
&\quad + \frac{1}{N_i^{2k}} E_\omega \int_0^t \Big\{ \sum_{\xi \neq \xi_i} g(X_\xi^{N_i}(s)) 1_{\{\|\xi\| \le k\}} 2d \Big\} ds.
\end{aligned} \tag{4.4.32}$$

Here

$$\begin{aligned}
&\frac{1}{N_i^{2k}} E_\omega \int_0^t \Big\{ \sum_{\xi \neq \xi_i, \alpha} \sum_{l=1}^\infty \Big(\frac{c}{N_i} \Big)^{l-1} \Big[X_\xi^{N_i, l, \alpha}(s) - X_\xi^{N_i, \alpha}(s) \Big] \\
&\qquad\qquad 1_{\{\|\xi\| \le k\}}\, 2 \sum_{\substack{\eta : \|\eta\| \le k \\ \eta \neq \eta_i}} \Big(X_\eta^{N_i, \alpha}(s) - x_\xi \Big) \Big\} ds \\
&\le E_\omega \int_0^t \Big\{ \sum_\alpha \sum_{l=1}^\infty \Big(\frac{c}{N_i} \Big)^{l-1} \frac{1}{N_i^k} \sum_{\xi : \|\xi\| \le k} \Big[X_\xi^{N_i, l, \alpha}(s) - X_\xi^{N_i, \alpha}(s) \Big] \\
&\qquad\qquad 2 \frac{1}{N_i^k} \sum_{\eta : \|\eta\| \le k} \Big(X_\eta^{N_i, \alpha}(s) - x_\xi \Big) \Big\} ds \\
&\quad + E_\omega \int_0^t \Big\{ \sum_{l=1}^\infty \Big(\frac{c}{N_i} \Big)^{l-1} \Big(\frac{2N_i^k - 1}{N_i^{2k}} \Big) R^2 \Big\} ds \\
&\le E_\omega \int_0^t \Big\{ \sum_\alpha \sum_{l=1}^\infty \Big(\frac{c}{N_i} \Big)^{l-1} \Big[X_\xi^{N_i, l \vee k, \alpha}(s) - X_\xi^{N_i, k, \alpha}(s) \Big] 2 \Big(X_\eta^{N_i, k, \alpha}(s) - x_\xi \Big) \Big\} ds \\
&\quad + t 2 \Big(\frac{2N_i^k - 1}{N_i^{2k}} R^2 \Big) \le 4t \Big(\frac{c}{N_i} \Big)^k R^2 + 4t \frac{1}{N_i^k} R^2 \le 8 R^2 \frac{t}{N_i^k},
\end{aligned} \tag{4.4.33}$$

where

$$\frac{1}{N_i^{2k}} E_\omega \int_0^t \Big\{ \sum_{\xi \neq \xi_i} g(X_\xi^{N_i}(s)) 1_{\{\|\xi\| \leq k\}} 2d \Big\} ds \leq 2d \|g\|_\infty \frac{t}{N_i^k}. \tag{4.4.34}$$

This completes the proof for $k \geq k_i$. By inspection we see that our bounds are also valid for $k < k_i$. ∎

Corollary 4.4.2 *For $i \in \mathbb{N}$ let X^{N_i} be a solution of (4.1.5) with initial condition (4.4.22) and let τ_i be as in Section 4.4.2. There exists a constant M such that for all $i \in \mathbb{N}$, $k \geq 1$, $x \in K^{\Omega_{N_i}}$ and $\omega \in \mathcal{C}_{\xi_i, K}[0, \infty)$*

$$E_\omega \left[\left| X_0^{N_i,k}(\tau_i) - x_0^k \right|^2 \right] \leq M \frac{\lambda_i}{N_i^k}. \tag{4.4.35}$$

Proof of Corollary 4.4.2: Condition on τ_i and use Lemma 4.4.1 to get

$$\begin{aligned} E_\omega \left[\left| X_0^{N_i,k}(\tau_i) - x_0^k \right|^2 \right] &= \int_0^\infty E_\omega \left[\left| X_0^{N_i,k}(t) - x_o^k \right|^2 \right] \lambda_i^{-1} e^{-t/\lambda_i} dt \\ &\leq \int_0^\infty M \frac{t}{N_i^k} \lambda_i^{-1} e^{-t/\lambda_i} dt = M \frac{\lambda_i}{N_i^k}. \end{aligned} \tag{4.4.36}$$

∎

4.4.4 An approximate equilibrium equation

Write

$$\Delta_i := \{\xi \in \Omega_{N_i} : \|\xi\| \leq k_i - 1\} \qquad (i \in \mathbb{N}) \tag{4.4.37}$$

for the $(k_i - 1)$-block around the origin in the hierarchical group Ω_{N_i}. For $\hat{\theta} \in K$ and $i \in \mathbb{N}$, we introduce an operator $A^i_{\hat{\theta}}$ with domain $\mathcal{D}(A^i_{\hat{\theta}}) := \mathcal{C}^2(K^{\Delta_i})$ and

$$\begin{aligned} &(A^i_{\hat{\theta}} f)(x) \\ &:= \sum_{\xi \in \Delta_i} \sum_{k=1}^{k_i-1} \left(\frac{c}{N_i} \right)^{k-1} \sum_\alpha [x_\xi^{k,\alpha} - x_\xi^\alpha] \tfrac{\partial}{\partial x_\xi^\alpha} f(x) + \sum_{\xi \in \Delta_i} g(x_\xi) \sum_\alpha (\tfrac{\partial}{\partial x_\xi^\alpha})^2 f(x) \\ &\quad + \sum_{\xi \in \Delta_i} \left(\frac{c}{N_i} \right)^{k_i-1} \sum_\alpha [\hat{\theta}^\alpha - x_\xi^\alpha] \tfrac{\partial}{\partial x_\xi^\alpha} f(x) \\ &= \sum_{\xi \in \Delta_i} \sum_{\eta \in \Delta_i} a^i(\eta - \xi) \sum_\alpha [x_\eta^\alpha - x_\xi^\alpha] \tfrac{\partial}{\partial x_\xi^\alpha} f(x) + \sum_{\xi \in \Delta_i} g(x_\xi) \sum_\alpha (\tfrac{\partial}{\partial x_\xi^\alpha})^2 f(x) \\ &\quad + \sum_{\xi \in \Delta_i} \left(\frac{c}{N_i} \right)^{k_i-1} \sum_\alpha [\hat{\theta}^\alpha - x_\xi^\alpha] \tfrac{\partial}{\partial x_\xi^\alpha} f(x), \end{aligned} \tag{4.4.38}$$

where

$$a^i(\xi) := \sum_{k=\|\xi\|}^{k_i-1} \frac{1}{N_i^k}\Big(\frac{c}{N_i}\Big)^{k-1} \qquad (\xi \in \Delta_i). \tag{4.4.39}$$

Lemma 4.4.3 below gives us an estimate of how close the expectation of functions f_i of large (k_i-1)-blocks is to their equilibrium value with respect to the dynamics described by the operator $A^i_{x_0^{k_i}}$.

Lemma 4.4.3 *For $i \in \mathbb{N}$ let X^{N_i} be a solution of (4.1.5) with initial condition (4.4.22) and let τ_i be as in Section 4.4.2. For $i \in \mathbb{N}$, let $f_i \in \mathcal{C}^2(K^{\Delta_i})$, let*

$$\nabla_\xi f_i := (\tfrac{\partial}{\partial x_\xi^\alpha} f_i)^{\alpha=1,\ldots,d}, \tag{4.4.40}$$

and let $\|\nabla_\xi f_i\|_\infty$ be the supremum norm of $|\nabla_\xi f_i|$ (the Euclidean norm of $\nabla_\xi f_i$). Then there exist constants M_1, M_2 such that for all $i \in \mathbb{N}$, $x \in K^{\Omega_{N_i}}$ and $\omega \in \mathcal{C}_{\xi_i,K}[0,\infty)$

$$\begin{aligned}&\Big|E_\omega\Big[(A^i_{x_0^{k_i}} f_i)(X^{N_i}(\tau_i))\Big]\Big| \\ &\qquad \leq 2\lambda_i^{-1}\|f_i\|_\infty + \Big\{M_1 c^{k_i-1} N_i^{1-\frac{3}{2}k_i}\lambda_i^{\frac{1}{2}} + M_2 c^{k_i} N_i^{-k_i}\Big\}\sum_{\xi\in\Delta_i}\|\nabla_\xi f_i\|_\infty.\end{aligned} \tag{4.4.41}$$

Here, in the left-hand side of (4.4.41), we lift the function $A^i_{x_0^k} f_i$ to the space $K^{\Omega_{N_i}}$ in the obvious way.

Proof of Lemma 4.4.3: Use the fact that the process M in (4.4.25) is a martingale and apply optional stopping to see that

$$E_\omega[f_i(X^{N_i}(\tau_i))] - E_\omega[f_i(X^{N_i}(0))] = E_\omega\Big[\int_0^{\tau_i} (Af_i)(X^{N_i}(t))\,dt\Big], \tag{4.4.42}$$

where for $x \in K^{\Omega_{N_i}}$

$$(Af_i)(x) = \sum_{\xi\in\Delta_i,\alpha}\sum_{k=1}^{\infty}\Big(\frac{c}{N}\Big)^{k-1}[x_\xi^{k,\alpha} - x_\xi^\alpha]\tfrac{\partial}{\partial x_\xi^\alpha} f_i(x) + \sum_{\xi\in\Delta_i,\alpha} g(x_\xi)(\tfrac{\partial}{\partial x_\xi^\alpha})^2 f_i(x), \tag{4.4.43}$$

and we have lifted the function f_i to $K^{\Omega_{N_i}}$ in the obvious way. Note that $\xi_i \notin \Delta_i$, so that in our summations we do not have to write $\xi \neq \xi_i$.

Formula (4.4.42) can be rewritten as

$$\begin{aligned}
\int_0^\infty dt\, E_\omega\Big[f_i(X^{N_i}(t))\Big]\lambda_i^{-1}e^{-t/\lambda_i} - E_\omega\Big[f_i(X^{N_i}(\beta_i t))\Big] \\
&= \int_0^\infty dt\; E_\omega\Big[\int_0^t du\,(Af_i)(X^{N_i}(u))\Big]\lambda_i^{-1}e^{-t/\lambda_i} \\
&= \int_0^\infty dt\; \int_0^t du\; E_\omega\Big[(Af_i)(X^{N_i}(u))\Big]\lambda_i^{-1}e^{-t/\lambda_i} \\
&= \Big\{(-e^{-t/\lambda_i})\int_0^t du\; E_\omega\Big[(Af_i)(X^{N_i}(u))\Big]\Big\}_{t=0}^{\infty} \\
&\quad - \int_0^\infty dt\;(-e^{-t/\lambda_i})E_\omega\Big[(Af_i)(X^{N_i}(t))\Big] \\
&= \int_0^\infty dt\; E_\omega\Big[(Af_i)(X^{N_i}(t))\Big]e^{-t/\lambda_i},
\end{aligned} \tag{4.4.44}$$

leading to the equation

$$E_\omega\Big[(Af_i)(X^{N_i}(\tau_i))\Big] = \lambda_i^{-1}\Big(E_\omega\Big[f_i(X^{N_i}(\tau_i))\Big] - E_\omega\Big[f_i(X^{N_i}(0))\Big]\Big). \tag{4.4.45}$$

Here

$$\begin{aligned}
&(Af_i)(X^{N_i}(\tau_i)) \\
&= \sum_{\xi\in\Delta_i}\sum_{k=1}^{\infty}\left(\frac{c}{N_i}\right)^{k-1}\sum_\alpha[X_\xi^{N_i,k,\alpha}(\tau_i) - X_\xi^{N_i,\alpha}(\tau_i)](\tfrac{\partial}{\partial x_\xi^\alpha}f_i)(X^{N_i}(\tau_i)) \\
&\quad + \sum_{\xi\in\Delta_i} g(X_\xi^{N_i}(\tau_i))\Big(\sum_\alpha(\tfrac{\partial}{\partial x_\xi^\alpha})^2 f_i\Big)(X^{N_i}(\tau_i)).
\end{aligned} \tag{4.4.46}$$

In view of our heuristic reasoning in Section 4.4.1, we write this as

$$\begin{aligned}
&(Af_i)(X^{N_i}(\tau_i)) \\
&= \sum_{\xi\in\Delta_i}\sum_{k=1}^{k_i-1}\left(\frac{c}{N_i}\right)^{k-1}\sum_\alpha[X_\xi^{N_i,k,\alpha}(\tau_i) - X_\xi^{N_i,\alpha}(\tau_i)](\tfrac{\partial}{\partial x_\xi^\alpha}f_i)(X^{N_i}(\tau_i)) \\
&\quad + \sum_{\xi\in\Delta_i} g(X_\xi^{N_i}(\tau_i))\Big(\sum_\alpha(\tfrac{\partial}{\partial x_\xi^\alpha})^2 f_i\Big)(X^{N_i}(\tau_i)) \\
&\quad + \sum_{\xi\in\Delta_i}\left(\frac{c}{N_i}\right)^{k_i-1}\sum_\alpha[x_0^{k_i,\alpha} - X_\xi^{N_i,\alpha}(\tau_i)](\tfrac{\partial}{\partial x_\xi^\alpha}f_i)(X^{N_i}(\tau_i)) \\
&\quad + \sum_{\xi\in\Delta_i}\left(\frac{c}{N_i}\right)^{k_i-1}\sum_\alpha[X_0^{N_i,k_i,\alpha}(\tau_i) - x_0^{k_i,\alpha}](\tfrac{\partial}{\partial x_\xi^\alpha}f_i)(X^{N_i}(\tau_i)) \\
&\quad + \sum_{\xi\in\Delta_i}\sum_{k=k_i+1}^{\infty}\left(\frac{c}{N_i}\right)^{k-1}\sum_\alpha[X_0^{N_i,k,\alpha}(\tau_i) - X_\xi^{N_i,\alpha}(\tau_i)](\tfrac{\partial}{\partial x_\xi^\alpha}f_i)(X^{N_i}(\tau_i)).
\end{aligned} \tag{4.4.47}$$

Here the first two terms represent the 'internal' evolution of the $(k_i - 1)$-block around the origin. The third term comes from their interaction with the k_i-block average, which for an appropriate choice of λ_i is essentially fixed to its value at time 0. The fourth term is an error term compensating for the fact that the k_i-block is not completely fixed at its value at time 0. The fifth term describes the (small) interaction with the k-blocks for $k > k_i$.

Combining (4.4.45) and (4.4.47), we arrive at

$$\begin{aligned}
E_\omega\Big[(A^i_{x_0^{k_i}} f_i)(X^{N_i}(\tau_i))\Big] &= \lambda_i^{-1}\Big(E_\omega\Big[f_i(X^{N_i}(\tau_i))\Big] - E_\omega\Big[f_i(X^{N_i}(\beta_i t))\Big]\Big) \\
&- \left(\frac{c}{N_i}\right)^{k_i-1} \sum_{\xi\in\Delta_i} E_\omega\Big[\sum_\alpha [X_0^{N_i,k_i,\alpha}(\tau_i) - x_0^{k_i,\alpha}](\tfrac{\partial}{\partial x_\xi^\alpha} f_i)(X^{N_i}(\tau_i))\Big] \\
&- \sum_{k=k_i+1}^{\infty} \left(\frac{c}{N_i}\right)^{k-1} \sum_{\xi\in\Delta_i} E_\omega\Big[\sum_\alpha [X_0^{N_i,k,\alpha}(\tau_i) - X_\xi^{N_i,\alpha}(\tau_i)](\tfrac{\partial}{\partial x_\xi^\alpha} f_i)(X^{N_i}(\tau_i))\Big].
\end{aligned} \tag{4.4.48}$$

From this it follows that

$$\begin{aligned}
\Big|E_\omega\Big[(A^i_{x_0^{k_i}} f_i)(X^{N_i}(\tau_i))\Big]\Big| &\le 2\lambda_i^{-1}\|f_i\|_\infty \\
&+ \left(\frac{c}{N_i}\right)^{k_i-1} E_\omega\Big[\Big|X_0^{N_i,k_i}(\tau_i) - x_0^{k_i}\Big|\Big] \sum_{\xi\in\Delta_i} \|\nabla_\xi f_i\|_\infty \\
&+ 2R\left(\frac{c}{N_i}\right)^{k_i} \sum_{\xi\in\Delta_i} \|\nabla_\xi f_i\|_\infty,
\end{aligned} \tag{4.4.49}$$

where we use that $N_i \ge 2$ and $c < 1$. We now note that by Corollary 4.4.2

$$E_\omega\Big[\Big|X_0^{N_i,k_i}(\tau_i) - x_0^{k_i}\Big|\Big]^2 \le E_\omega\Big[\Big|X_0^{N,k}(\tau_i) - x_0^{k_i}\Big|^2\Big] \le M\lambda_i N_i^{-k_i}. \tag{4.4.50}$$

Inserting this into (4.4.49), we arrive at (4.4.41). ∎

4.4.5 Equilibrium calculations

In this section, we construct 'test functions' f_i that we will insert into the almost-equilibrium equation in Lemma 4.4.3 to draw certain conclusions about the behavior of the $(k_i - 1)$-blocks.

We split the operator $A^i_{\hat\theta}$ in (4.4.38) as follows:

$$A^i_{\hat\theta} = B^i_{\hat\theta} + C^i, \tag{4.4.51}$$

where for $f \in \mathcal{C}^2(K^{\Delta_i})$

$$
\begin{aligned}
(B^i_{\hat\theta} f)(x) &:= \sum_{\xi\in\Delta_i}\sum_{\eta\in\Delta_i} a^i(\eta-\xi)\sum_\alpha [x^{k,\alpha}_\xi - x^\alpha_\xi]\tfrac{\partial}{\partial x^\alpha_\xi} f(x) \\
&\quad + \sum_{\xi\in\Delta_i}\left(\frac{c}{N_i}\right)^{k_i-1}\sum_\alpha [\hat\theta^\alpha - x^\alpha_\xi]\tfrac{\partial}{\partial x^\alpha_\xi} f(x) \\
(C^i f)(x) &:= \sum_{\xi\in\Delta_i} g(x_\xi)\sum_\alpha (\tfrac{\partial}{\partial x^\alpha_\xi})^2 f(x).
\end{aligned}
\tag{4.4.52}
$$

A central role will be played by the semigroup generated by (an extension of) $B^i_{\hat\theta}$. There is an explicit formula for this semigroup, in terms of a continuous-time random walk on Δ_i that jumps from site ξ to site η with rate $a^i(\eta-\xi)$ and that is killed in each site with rate $(\frac{c}{N_i})^{k_i-1}$. Let $P^i_t(\eta-\xi)$ be the probability that this random walk, starting in ξ, is in η at time t. For each $i\in\mathbb{N}$, $\hat\theta\in K$ and $t\geq 0$ put

$$(P^i_{\hat\theta,t}x)_\xi := \hat\theta + \sum_{\eta\in\Delta_i} P^i_t(\xi-\eta)[x_\eta - \hat\theta], \tag{4.4.53}$$

and let $(R^i_{\hat\theta,t})_{t\geq 0}$ be the Feller semigroup on $\mathcal{C}(K^{\Delta_i})$ given by

$$(R^i_{\hat\theta,t} f)(x) := f(P^i_t x). \tag{4.4.54}$$

The generator of $(R^i_{\hat\theta,t})_{t\geq 0}$ is an extension of the operator $B^i_{\hat\theta}$. In particular, for any $f\in\mathcal{C}^2(K^{\Delta_i})$ one has

$$\tfrac{\partial}{\partial t}(R^i_t f)(x) = (B^i_{\hat\theta} R^i_t f)(x) \qquad (x\in K^{\Delta_i}). \tag{4.4.55}$$

With these definitions, we construct our test functions f_i as follows:

$$f_i(x) := \int_0^\infty \left\{ g^*(\hat\theta) - g^*\Big(\hat\theta + \sum_\eta P^i_t(0-\eta)[x_\eta - \hat\theta]\Big)\right\} dt \qquad (x\in K^{\Delta_i}). \tag{4.4.56}$$

Inserting these test functions into the almost-equilibrium equation in Lemma 4.4.3, we arrive at the following result.

Lemma 4.4.4 *For $i\in\mathbb{N}$ let X^{N_i} be a solution of (4.1.5) with initial condition (4.4.22) and let τ_i be as in Section 4.4.2. Then there exist constants M_1, M_2, M_3 such that for all $i\in\mathbb{N}$, $x\in K^{\Omega_{N_i}}$ and $\omega\in\mathcal{C}_{\xi_i,K}[0,\infty)$*

$$
\begin{aligned}
&\Big| g^*(x_0^{k_i}) - E_\omega\Big[g^*(X_0^{N_i}(\tau_i))\Big] - \mu_i E_\omega\Big[g(X_0^{N_i}(\tau_i))\Big]\Big| \\
&\qquad\leq M_1\lambda_i^{-1}c^{1-k_i}N_i^{k_i-1} + M_2\lambda_i^{\frac12}N_i^{-\frac12 k_i} + M_3 c N_i^{-1},
\end{aligned}
\tag{4.4.57}
$$

where

$$\mu_i := 2\int_0^\infty \sum_{\xi\in\Delta_i} P_t^i(\xi)^2 dt. \tag{4.4.58}$$

$P_t^i(\xi)$ is defined just before formula (4.4.53). The constants M_1 and M_2 here are not the same constants as in Lemma 4.4.3.

Proof of Lemma 4.4.4: Put $\hat\theta := x_0^{k_i}$ and define a function $g^*_{\hat\theta} \in \mathcal{C}(K^{\Delta_i})$ by

$$g^*_{\hat\theta}(x) := g^*(x_0) - g^*(\hat\theta) \qquad (x \in K^{\Delta_i}). \tag{4.4.59}$$

Then the test function f_i can be written as

$$f_i(x) = \int_0^\infty \Big\{(R^i_{\hat\theta,t} g^*_{\hat\theta})(x)\Big\} dt \qquad (x \in K^{\Delta_i}) \tag{4.4.60}$$

where $(R^i_{\hat\theta,t})_{t\geq 0}$ is the semigroup in (4.4.54). Our first aim is now to show that

$$(A^i_{\hat\theta} f_i)(x) = g^*(\hat\theta) - g^*(x_0) - 2\int_0^\infty \sum_{\xi\in\Delta_i} P_t^i(\xi)^2 g(x_\xi) dt. \tag{4.4.61}$$

To see that (4.4.61) holds, write

$$\begin{aligned}
(A^i_{\hat\theta} f_i)(x) &= \int_0^\infty \Big\{(B^i_{\hat\theta} R^i_{\hat\theta,t} g^*_{\hat\theta})(x) + (C^i R^i_{\hat\theta,t} g^*_{\hat\theta})(x)\Big\} dt \\
&= \int_0^\infty \Big\{\tfrac{\partial}{\partial t}(R^i_{\hat\theta,t} g^*_{\hat\theta})(x) + (C^i R^i_{\hat\theta,t} g^*_{\hat\theta})(x)\Big\} dt \\
&= (R^i_{\hat\theta,\infty} g^*_{\hat\theta})(x) - (R^i_{\hat\theta,0} g^*_{\hat\theta})(x) + \int_0^\infty \Big\{(C^i R^i_{\hat\theta,t} g^*_{\hat\theta})(x)\Big\} dt \\
&= g^*(\hat\theta) - g^*(x_0) + \int_0^\infty \Big\{(C^i R^i_{\hat\theta,t} g^*_{\hat\theta})(x)\Big\} dt.
\end{aligned} \tag{4.4.62}$$

In order to evaluate the third term, split the function $g^*_{\hat\theta}$ as

$$g^*_{\hat\theta}(x) = h_{\hat\theta}(x) - f_{\hat\theta}(x) \qquad (x \in K^{\Delta_i}), \tag{4.4.63}$$

where

$$\left.\begin{aligned} f_{\hat\theta}(x) &:= \tfrac{1}{d}|x_0 - \hat\theta|^2 \\ h_{\hat\theta}(x) &= h(x_0) \end{aligned}\right\} \qquad (x \in K^{\Delta_i}), \tag{4.4.64}$$

with $h \in \mathcal{C}^2(K^\circ) \cap \mathcal{C}(K)$ a harmonic function, i.e.

$$\Delta h(x) = 0 \qquad (x \in K^\circ). \tag{4.4.65}$$

A straightforward generalization of Lemma 1.6 in Swart [41] (Lemma 3.1.6 in this dissertation) gives

$$(C^i R^i_{\hat\theta,t} h_{\hat\theta})(x) = 0 \qquad (x \in K^{\Delta_i},\ t \geq 0), \tag{4.4.66}$$

while a straightforward calculation gives

$$\begin{aligned}(C^i R^i_t f_{\hat\theta})(x) &= \sum_{\xi\in\Delta_i} g(x_\xi) \sum_\alpha \left(\tfrac{\partial}{\partial x^\alpha_\xi}\right)^2 \tfrac{1}{d} \Big|\hat\theta + \sum_{\eta\in\Delta_i} P^i_t(0-\eta)[x_\eta - \hat\theta]\Big|^2 \\ &= 2\sum_{\xi\in\Delta_i} P^i_t(\xi)^2 g(x_\xi) \qquad\qquad (x \in K^{\Delta_i},\ t \geq 0)\end{aligned} \tag{4.4.67}$$

Here we used that $P^i_t(-\xi) = P^i_t(\xi)$. Inserting (4.4.66) and (4.4.67) into (4.4.62) we arrive at (4.4.61).

We now want to insert the test functions f_i into the almost equilibrium equation in Lemma 4.4.3. For this, we need bounds on $\|f_i\|_\infty$ and $\|\nabla_\xi f_i\|_\infty$. We have the following:

$$\begin{aligned}\|f_i\|_\infty &\leq R\|\nabla g^*\|_\infty \int_0^\infty \sum_{\xi\in\Omega^i} P^i_t(\xi)dt \\ \sum_{\xi\in\Omega^i} \|\nabla_\xi f_i\|_\infty &\leq \|\nabla g^*\|_\infty \int_0^\infty \sum_{\xi\in\Omega^i} P^i_t(\xi)dt,\end{aligned} \tag{4.4.68}$$

where

$$\nabla g^* := \left(\tfrac{\partial}{\partial x^\alpha} g^*\right)^{\alpha=1,\dots,d} \tag{4.4.69}$$

and $\|\nabla g^*\|_\infty$ is the supremum norm of $|\nabla g^*|$. To see that the estimate for $\|f_i\|_\infty$ in (4.4.68) holds, note that

$$\Big|\sum_{\eta\in\Omega^i} P^i_t(0-\eta)[x_\eta - \hat\theta]\Big| \leq R\sum_{\xi\in\Omega^i} P^i_t(\xi), \tag{4.4.70}$$

so that

$$\left|g^*(\hat\theta) - g^*\Big(\hat\theta + \sum_\eta P^i_t(0-\eta)[x_\eta - \hat\theta]\Big)\right| \leq \|\nabla g^*\|_\infty R \sum_{\xi\in\Omega^i} P^i_t(\xi). \tag{4.4.71}$$

To see that the estimate for $\|\nabla_\xi f_i\|_\infty$ holds, simply note that

$$\begin{aligned}&\tfrac{\partial}{\partial x^\alpha_\xi}\left\{g^*(\hat\theta) - g^*\Big(\hat\theta + \textstyle\sum_\eta P^i_t(0-\eta)[x_\eta - \hat\theta]\Big)\right\} \\ &= P^i_t(\xi)\left(\tfrac{\partial}{\partial x^\alpha} g^*\right)\Big(\hat\theta + \textstyle\sum_\eta P^i_t(0-\eta)[x_\eta - \hat\theta]\Big).\end{aligned} \tag{4.4.72}$$

Next, we use that

$$\int_0^\infty dt \sum_{\xi \in \Omega^i} P_t^i(\xi) = \left(\frac{N_i}{c}\right)^{k_i - 1}. \tag{4.4.73}$$

This is just the expected survival time of a random walk that is killed in every site with rate $(\frac{c}{N_i})^{k_i-1}$. Inserting (4.4.73) into (4.4.68), and inserting (4.4.61) into the almost-equilibrium equation in Lemma 4.4.3, we find that

$$\begin{aligned}\left|E_\omega\left[g^*(\hat{\theta}) - g^*(X_0^{N_i}(\tau_i)) - 2\int_0^\infty \sum_{\xi \in \Delta_i} P_t^i(\xi)^2 g(X_\xi^{N_i}(\tau_i))dt\right|\right| \\ \leq 2\lambda_i^{-1} R \|\nabla g^*\|_\infty \left(\tfrac{N_i}{c}\right)^{k_i-1} \\ + \left\{M_1 c^{k_i-1} N_i^{1-\frac{3}{2}k_i} \lambda_i^{\frac{1}{2}} + M_2 c^{k_i} N_i^{-k_i}\right\} \|\nabla g^*\|_\infty \left(\tfrac{N_i}{c}\right)^{k_i-1}.\end{aligned} \tag{4.4.74}$$

Using the fact that by the symmetry between all sites in the $(k_i - 1)$-block around the origin

$$E_\omega[g(X_\xi^{N_i}(\tau_i))] = E_\omega[g(X_0^{N_i}(\tau_i))] \qquad (\|\xi\| \leq k_i - 1) \tag{4.4.75}$$

we arrive at (4.4.57). ∎

4.4.6 Asymptotics of the scaling factor μ_i

In this section we investigate the asymptotics as $i \to \infty$ of the quantity μ_i defined in (4.4.58). In particular, we prove the following.

Lemma 4.4.5

$$\mu_{k_i} \sim \sigma_{k_i} \qquad \text{as } i \to \infty. \tag{4.4.76}$$

Proof of Lemma 4.4.5: Consider the $(k-1)$-block

$$\Delta(N, k) := \{\xi \in \Omega_N : \|\xi\| \leq k - 1\} \qquad (N \geq 2,\ k \geq 1), \tag{4.4.77}$$

and put

$$a_N^k(\xi) := \sum_{l=\|\xi\|}^{k-1} \frac{1}{N^l}\left(\frac{c}{N}\right)^{l-1}. \tag{4.4.78}$$

Let $\{(N_t^l)_{t\geq 0}\}_{l=1,2,\ldots}$ be a collection of independent (right-continuous) Poisson processes, where the process $(N_t^l)_{t\geq 0}$ has intensity $(\frac{c}{N})^{l-1}$. Let $(\xi_n^l)_{n=1,2,\ldots}^{l=1,2,\ldots}$ be a collection of independent Ω_N-valued random variables with

$$P[\xi_n^l = \xi] = \frac{1}{N^l} 1_{\{\|\xi\| \leq l\}} \qquad (\xi \in \Omega_N). \tag{4.4.79}$$

Then the process $(I_t^k)_{t\geq 0}$, defined as

$$I_t^k := \sum_{l=1}^{k-1} \sum_{n=1}^{N_t^l} \xi_n^l, \tag{4.4.80}$$

performs a random walk on $\Delta(N,k)$ that jumps from a site ξ to a site η with rate $a_N^k(\eta-\xi)$. (Recall that summation in Ω_N is defined pointwise modulo N.) For $l = 1, 2, \ldots$ we introduce stopping times

$$\tau_l := \min\{t \geq 0 : N_t^l > 0\} \qquad (l = 1, 2, \ldots), \tag{4.4.81}$$

so that $\mu_i = \mu(N_i, k_i)$, with

$$\mu(N,k) := 2\int_0^\infty \sum_{\xi\in\Delta(N,k)} P[I_t^k \in \xi,\ t < \tau_k]^2 dt. \tag{4.4.82}$$

Here $P[I_t^k \in \xi,\ t \leq \tau_k]^2$ is the probability that two independent copies of the process are both in ξ. For such independent copies we may write

$$\mu(N,k) = 2\int_0^\infty P[I_t^k = \tilde{I}_t^k,\ t < \tau_k \wedge \tilde{\tau}_k] dt. \tag{4.4.83}$$

Next, we use the fact that the difference process $(I_t^i - \tilde{I}_t^i)_{t\geq 0}$ performs a random walk on Δ_i that jumps from a site ξ to a site η with rate $2a^i(\eta-\xi)$, while $\tau_k \wedge \tilde{\tau}_k$ is exponentially distributed with mean $\frac{1}{2}E[\tau_k]$. Absorbing the factor 2 in a redefinition of the time scale we see that

$$\mu(N,k) = \int_0^\infty P[I_t^k = 0,\ t \leq \tau_k] dt = E\int_0^{\tau_k} 1_{\{I_t^k = 0\}} dt. \tag{4.4.84}$$

We put $\tau_0 := 0$ and introduce stopping times

$$\sigma_l^k := \min\{\tau_l, \ldots, \tau_k\} \qquad (l = 0, \ldots, k). \tag{4.4.85}$$

We split (4.4.84) as

$$\mu(N,k) = \sum_{l=0}^{k-1} E\int_{\sigma_l^k}^{\sigma_{l+1}^k} 1_{\{I_t^k = 0\}} dt. \tag{4.4.86}$$

By symmetry,

$$P[I_t^k = \xi | \sigma_l^k \leq t < \sigma_{l+1}^k] = 1_{\{\|\xi\| \leq l\}} \frac{1}{N^l} \qquad (l = 0, \ldots, k-1), \tag{4.4.87}$$

and hence

$$\mu(N,k) = \sum_{l=0}^{k-1} \frac{1}{N^l} E[\sigma_{l+1}^k - \sigma_l^k], \tag{4.4.88}$$

where

$$E[\sigma_{l+1}^k - \sigma_l^k] = \left(\frac{N}{c}\right)^l P[\sigma_l^k = \tau_l]. \tag{4.4.89}$$

We therefore find that

$$\mu(N,k) = \sum_{l=0}^{k-1} c^{-l} P[\sigma_l^k = \tau_l]. \tag{4.4.90}$$

Here $P[\sigma_0^k = \tau_0] = 1$ and for $l \geq 1$

$$\begin{aligned} P[\sigma_l^k < \tau_l] &= P[\tau_n < \tau_l \text{ for some } n = l+1, \dots, k] \\ &\leq P[\tau_n < \tau_l \text{ for some } n > l] \\ &= \frac{\sum_{n=l+1}^{\infty} (\frac{c}{N})^n}{\sum_{n=l}^{\infty} (\frac{c}{N})^n} = \frac{c}{N}, \end{aligned} \tag{4.4.91}$$

where we used that for independent exponentially distributed variables X_1, X_2 with mean $\lambda_1^{-1}, \lambda_2^{-1}$:

$$P[X_1 < X_2] = \frac{\lambda_1}{\lambda_1 + \lambda_2}. \tag{4.4.92}$$

Inserting (4.4.91) into (4.4.90) we see that

$$\left(1 - \frac{c}{N}\right) \sum_{l=0}^{k-1} c^{-l} \leq \mu(N,k) \leq \sum_{l=0}^{k-1} c^{-l}, \tag{4.4.93}$$

and using the fact that $N_i \to \infty$ as $i \to \infty$, we arrive at (4.4.76). ∎

4.4.7 Proof of the convergence

Lemma 4.4.6 *For $i \in \mathbb{N}$ let X^{N_i} be a solution of (4.1.5) with initial condition (4.4.22) and let τ_i be as in Section 4.4.2. Assume that*

$$c^{1-k_i} N_i^{k_i - 1} \ll \lambda_i \ll N_i^{k_i}. \tag{4.4.94}$$

Then there exist constants M_i, tending to zero as $i \to \infty$, such that for all $x \in K^{\Omega_{N_i}}$ and $\omega \in \mathcal{C}_{\xi_i,K}[0,\infty)$

$$\left| g^*(x_0^{k_i}) - \sigma_{k_i} E_\omega\Big[g(X_0^{N_i}(\tau_i))\Big] \right| \leq M_i. \tag{4.4.95}$$

Note that condition (4.4.94) can be met because of condition (4.1.27).

Proof of Lemma 4.4.6: Combining Lemma 4.4.4, Lemma 4.4.5 and formula (4.4.94) we see that there exist constants M_i, tending to zero as $i \to \infty$, such that for all $x \in K^{\Omega_{N_i}}$ and $\omega \in \mathcal{C}_{\xi_i,K}[0,\infty)$

$$\left| g^*(x_0^{k_i}) - E_\omega\Big[g^*(X_0^{N_i}(\tau_i))\Big] - \sigma_{k_i} E_\omega\Big[g(X_0^{N_i}(\tau_i))\Big]\right| \le M_i. \tag{4.4.96}$$

We may conclude from (4.4.96) that

$$\left| E_\omega\Big[g(X_0^{N_i}(\tau_i))\Big]\right| \le (\sigma_{k_i})^{-1}(M_i + \|g^*\|_\infty). \tag{4.4.97}$$

Since $\sigma_{k_i} \to \infty$ as $i \to \infty$ and since g is continuous on K and nonzero on K°, there exists constants $\tilde{M}_i$, tending to zero as $i \to \infty$, such that for all $x \in K^{\Omega_{N_i}}$ and $\omega \in \mathcal{C}_{\xi_i,K}[0,\infty)$

$$\left| E_\omega\Big[g^*(X_0^{N_i}(\tau_i))\Big]\right| \le \tilde{M}_i. \tag{4.4.98}$$

When we insert this into (4.4.96), then after a suitable redefinition of our constants M_i we arrive at (4.4.95). ∎

We now translate the statement in Lemma 4.4.6 about the conditional law of X^{N_i} given

$$\left(X_{\xi_i}^{N_i}(t)\right)_{t\ge 0} = (\omega(t))_{t\ge 0} \tag{4.4.99}$$

into a statement about the unconditional law.

Lemma 4.4.7 *For $i \in \mathbb{N}$ let X^{N_i} be a solution of (4.1.5) with initial condition (4.4.22) and let λ_i ($i \in \mathbb{N}$) be constants satisfying (4.4.94). For $i \in \mathbb{N}$ and $\xi \in \Omega_{N_i}$ let Y_ξ^i be given by*

$$Y_\xi^i := \int_0^\infty \left(\sigma_{k_i} g(X_0^{N_i}(t)) - g^*(x_0^{k_i})\right)\lambda_i^{-1} e^{t/\lambda_i}\, dt. \tag{4.4.100}$$

Then there exist constants M_i, tending to zero as $i \to \infty$, such that for all $x \in K^{\Omega_{N_i}}$

$$\left| E\Big[Y_0^i Y_{\xi_i}^i\Big]\right| \le M_i. \tag{4.4.101}$$

Proof of Lemma 4.4.7: Note that almost surely $Y^i_\xi \geq -\|g^*\|_\infty$, and hence

$$\begin{aligned} E\Big[|Y^i_\xi|\Big] &= \int_{-\|g\|_\infty}^{\infty} yP[Y^i_\xi \in dy] \\ &= \int_{-\|g\|_\infty}^{0} yP[Y^i_\xi \in dy] + \int_0^\infty yP[Y^i_\xi \in dy] \\ &\leq \|g\|_\infty + E[Y^i_\xi] \qquad\qquad (i \in \mathbb{N},\ \xi \in \Omega_{N_i}). \end{aligned} \tag{4.4.102}$$

Lemma 4.4.6 implies that for a suitable version of the conditional expectation

$$|E[Y^i_0 | Y^i_{\xi_i} = y]| \leq M_i \qquad\qquad (y \geq -\|g^*\|_\infty), \tag{4.4.103}$$

and by symmetry the same is true for the conditional expectation of $Y^i_{\xi_i}$ given Y^i_0. It follows that

$$\begin{aligned} \Big|E[Y^i_0 Y^i_{\xi_i}]\Big| &= \Big|\int_{-\|g^*\|_\infty}^{\infty} E[Y^i_0 Y^i_{\xi_i} | Y^i_{\xi_i} = y] P[Y^i_{\xi_i} \in dy]\Big| \\ &= \Big|\int_{-\|g^*\|_\infty}^{\infty} yE[Y^i_0 | Y^i_{\xi_i} = y] P[Y^i_{\xi_i} \in dy]\Big| \\ &\leq \int_{-\|g^*\|_\infty}^{\infty} \Big|yE[Y^i_0 | Y^i_{\xi_i} = y]\Big| P[Y^i_{\xi_i} \in dy] \\ &\leq M_i(\|g^*\|_\infty + E[Y^i_{\xi_i}]) \leq M_i(\|g\|_\infty + M_i), \end{aligned} \tag{4.4.104}$$

where in the last step we used that

$$\begin{aligned} \Big|E[Y^i_{\xi_i}]\Big| &= \Big|\int_{-\|g^*\|_\infty}^{\infty} E[Y^i_{\xi_i} | Y^i_0 = y] P[Y^i_0 \in dy]\Big| \\ &\leq \int_{-\|g^*\|_\infty}^{\infty} \Big|E[Y^i_{\xi_i} | Y^i_0 = y]\Big| P[Y^i_0 \in dy] \leq M_i. \end{aligned} \tag{4.4.105}$$

∎

Lemma 4.4.8 *For $i \in \mathbb{N}$ let X^{N_i} be a solution of (4.1.5) with initial condition (4.4.22), and let λ_i ($i \in \mathbb{N}$) be constants satisfying (4.4.94). Then there exist constants M_i, tending to zero as $i \to \infty$, such that for all $x \in K^{\Omega_{N_i}}$*

$$E\Bigg[\Big|\int_0^\infty \Big(g^*(X^{N_i,k_i}_0(t)) - \sigma_{k_i}\frac{1}{N_i^{k_i}} \sum_{\xi:\|\xi\|\leq k_i} g(X^{N_i}_\xi(t))\Big)\lambda_i^{-1} e^{t/\lambda_i} dt\Big|^2\Bigg] \leq M_i. \tag{4.4.106}$$

Proof of Lemma 4.4.8: Defining random variables Y^i_ξ as in Lemma 4.4.7 and using symmetry, we see that

$$\begin{aligned}
&E\Bigg[\Big|\int_0^\infty \Big(g^*(x_0^{k_i}) - \sigma_{k_i}\frac{1}{N_i^{k_i}}\sum_{\xi:\|\xi\|\le k_i} g(X_\xi^{N_i}(t))\Big)\lambda_i^{-1}e^{t/\lambda_i}dt\Big|^2\Bigg] \\
&\quad = \frac{1}{N_i^{2k_i}}\sum_{\substack{\xi:\|\xi\|\le k_i\\ \eta:\|\eta\|\le k_i}} E\Big[Y^i_\xi Y^i_\eta\Big] \\
&\quad = \frac{N_i^{k_i}(N_i^{k_i}-N_i^{k_i-1})}{N_i^{2k_i}}E[Y^i_0 Y^i_{\xi_i}] + \frac{1}{N_i^{2k_i}}\sum_{\substack{\xi:\|\xi\|\le k_i\\ \eta:\|\eta\|\le k_i\\ \|\xi-\eta\|\le k_i-1}} E\Big[Y^i_\xi Y^i_\eta\Big] \\
&\quad \le E[Y^i_0 Y^i_{\xi_i}] + \frac{1}{N_i}\Big(\|g^*\|_\infty + \sigma_{k_i}\|g\|_\infty\Big),
\end{aligned} \tag{4.4.107}$$

where $\sigma_{k_i} \sim c^{-k_i}(c/(1-c))$ and hence $\sigma_{k_i}/N_i \to 0$ by (4.4.94). But

$$\begin{aligned}
&E\Bigg[\Big|\int_0^\infty \Big(g^*(x_0^{k_i}) - g^*(X_0^{N_i,k_i}(t))\Big)\lambda_i^{-1}e^{t/\lambda_i}dt\Big|^2\Bigg] \\
&\qquad \le E\Bigg[\int_0^\infty \Big|g^*(x_0^{k_i}) - g^*(X_0^{N_i,k_i}(t))\Big|^2\lambda_i^{-1}e^{t/\lambda_i}dt\Bigg] \\
&\qquad \le \|g^*\|_\infty E\Bigg[\int_0^\infty \Big|x_0^{k_i} - X_0^{N_i,k_i}(t)\Big|^2\lambda_i^{-1}e^{t/\lambda_i}dt\Bigg] \\
&\qquad \le \|g^*\|_\infty M\frac{\lambda_i}{N_i^{k_i}}
\end{aligned} \tag{4.4.108}$$

by Corollary 4.4.2. Here $\lambda_i/N_i^{k_i} \to 0$ as $i \to \infty$ by (4.4.94), and combining (4.4.107) and (4.4.108) and applying Lemma 4.4.7 we arrive at (4.4.106). ∎

Lemma 4.4.9 *For $i \in \mathbb{N}$ let X^{N_i} be a solution of (4.1.5) with initial condition (4.1.6). Then for $i \in \mathbb{N}$ there exists positive constants γ_i, M_i satisfying $\gamma_i \ll c^{k_i}$ and $M_i \ll 1$ ($i \to \infty$), such that for all $t \ge 0$*

$$E\Bigg[\Big|\int_0^\infty \Big(g^*(\hat{X}^i(t+s)) - \hat{G}^i(t+s)\Big)\gamma_i^{-1}e^{-s/\gamma_i}ds\Big|^2\Bigg] \le M_i. \tag{4.4.109}$$

Proof of Lemma 4.4.9: Let us write

$$\begin{aligned}
R_i(t) &:= g^*(\hat{X}^i(t)) - \hat{G}^i(t) \\
&= g^*(X_0^{N_i,k_i}(\beta_i t)) - \sigma_{k_i}\frac{1}{N_i^{k_i}}\sum_{\xi:\|\xi\|\le k_i} g(X_\xi^{N_i}(\beta_i t)).
\end{aligned} \tag{4.4.110}$$

Then

$$
\begin{aligned}
&E\left[\left|\int_0^\infty R_i(t+s/\beta_i)\,\lambda_i^{-1}e^{s/\lambda_i}ds\right|^2\right]\\
&=\int_{K^{\Omega_{N_i}}} E\left[\left|\int_0^\infty R_i(t+s/\beta_i)\,\lambda_i^{-1}e^{s/\lambda_i}ds\right|^2\;\middle|\; X^{N_i}(\beta_i t)=x\right]P\left[X^{N_i}(\beta_i t)\in dx\right]\\
&\le M_i,
\end{aligned}
\tag{4.4.111}
$$

where we applied Lemma 4.4.8 to the conditioned system. A simple change of variables now gives (4.4.109), where

$$
\gamma_i := \frac{\lambda_i}{\beta_i} \ll \frac{N_i^{k_i}}{\sigma_{k_i} N_i^{k_i}} \sim c^{k_i} c^{-1}(1-c). \tag{4.4.112}
$$

∎

Proof of formula (4.1.30): For $i \in \mathbb{N}$ let R_i be as in (4.4.110) and let τ_i be an exponentially distributed random variable with mean γ_i, independent of the process R_i. For any square integrable random variable X we write

$$
\|X\|_2 := E\left[|X|^2\right]. \tag{4.4.113}
$$

Then by Lemma 4.4.9, for each $T \ge 0$ and $\gamma > 0$ and for each i such that $\gamma_i < \gamma$:

$$
\begin{aligned}
&E\left[\left|\int_0^\infty R_i(t)1_{[T,\infty)}(t)\gamma^{-1}e^{(t-T)/\gamma}dt\right|^2\right]^{\frac{1}{2}}\\
&\quad= \|R_i(T+\tau_i+\sigma_i)\|_2\\
&\quad= \left\|\int_0^\infty R_i(T+\tau_i+s)(\gamma-\gamma_i)^{-1}e^{s/(\gamma-\gamma_i)}ds\right\|_2\\
&\quad\le \int_0^\infty \|R_i(T+\tau_i+s)\|_2\,(\gamma-\gamma_i)^{-1}e^{s/(\gamma-\gamma_i)}ds \le M_i^{\frac{1}{2}}.
\end{aligned}
\tag{4.4.114}
$$

where σ_i is an exponentially distributed random variable with mean $\gamma-\gamma_i$, independent of the process R_i and of τ_i. Let us write

$$
f_{T,\gamma}(t) := 1_{[T,\infty)}(t)\gamma^{-1}e^{(t-T)/\gamma}. \tag{4.4.115}
$$

Then (4.4.114) implies that for all $T \ge 0$ and $\gamma > 0$

$$
\lim_{i\to\infty} E\left[\left|\int_0^\infty R_i(t)f_{T,\gamma}(t)dt\right|^2\right] = 0. \tag{4.4.116}
$$

Let $\mathcal{G}$ be the class of measurable functions $f : [0, \infty) \to [0, \infty)$ such that

$$\lim_{i\to\infty} E\Big[\Big|\int_0^\infty R_i(t) f(t)dt\Big|^2\Big] = 0. \tag{4.4.117}$$

Then clearly

$$\begin{aligned} f, g \in \mathcal{G},\ f \geq g &\Rightarrow f - g \in \mathcal{G} \\ f \in \mathcal{G},\ \lambda \geq 0 &\Rightarrow \lambda f \in \mathcal{G} \\ f_i \in \mathcal{G}\ (i \in \mathbb{N}),\ f_i \downarrow f &\Rightarrow f \in \mathcal{G}. \end{aligned} \tag{4.4.118}$$

It follows that

$$1_{[0,T)} = \lim_{n\to\infty} \Big(e^{T/n} f_{0,n} - f_{T,n}\Big) \in \mathcal{G}, \tag{4.4.119}$$

and this proves formula (4.1.30). ∎

4.5 Discussion

Here we discuss why formula (4.1.30) is not sufficient to conclude that the $\hat{X}^i$ ($i \in \mathbb{N}$) are tight, i.e., that the collection of their laws is relatively compact in $\mathcal{P}(\mathcal{D}_K[0, \infty))$, the space of probability measures on $\mathcal{D}_K[0, \infty)$ equipped with the topology of weak convergence. (We remind the reader that $\mathcal{D}_K[0, \infty)$ is the space of cadlag functions from $[0, \infty)$ to K, equipped with the Skorohod topology.) The following is a special case of Theorem 9.4 from Chapter 3 of Ethier & Kurtz [16].

Proposition 4.5.1 *Let E be a compact metric space and let $\mathcal{D} \subset \mathcal{C}(E)$ be dense. For $n = 1, 2, \ldots$, let X_n be a process with sample paths in $\mathcal{D}_E[0, \infty)$ defined on a probability space $(\Omega_n, \mathcal{F}^n, P_n)$ and adapted to a filtration $(\mathcal{F}^n_t)_{t\geq 0}$. For each $f \in \mathcal{D}$ and $n = 1, 2, \ldots$ let A^f_n be an $(\mathcal{F}^n_t)_{t\geq 0}$-adapted process with sample paths in $\mathcal{D}_{\mathbb{R}}[0, \infty)$ and $\sup_{t\geq 0} E[|A^f_n(t)|] < \infty$. Assume that for each $f \in \mathcal{D}$ and $n = 1, 2, \ldots$ the process $(M(t))_{t\geq 0}$ given by*

$$M(t) := f(X_n(t)) - \int_0^t A^f_n(s)ds \qquad (t \geq 0) \tag{4.5.1}$$

is an $(\mathcal{F}^n_t)_{t\geq 0}$-martingale. Assume that for every $f \in \mathcal{D}$ there exists a $p \in (1, \infty)$ such that for all $T > 0$

$$\sup_{n=1,2,\ldots} E\Big[\Big(\int_0^T |A^f_n(t)|^p\ dt\Big)^{\frac{1}{p}}\Big] < \infty. \tag{4.5.2}$$

Then the sequence of k-dimensional processes $(f_1(X_n), \ldots, f_k(X_n))_{n=1,2,\ldots}$ is tight in path space $\mathcal{D}_{\mathbb{R}^k}[0, \infty)$ for every $f_1, \ldots, f_k \in \mathcal{C}(E)$.

Combining this with (4.1.29) and (4.1.28) it is not hard to see that tightness of $(\hat{X}^i)_{i\in\mathbb{N}}$ would follow if we could prove formula (4.1.31), i.e., if we would have

$$\lim_{i\to\infty} E\Big[\int_0^T \Big|\hat{G}^i(t) - g^*(\hat{X}^i(t))\Big|^2 dt\Big] = 0. \tag{4.5.3}$$

However, we have only shown that

$$\lim_{i\to\infty} E\left[\Big|\int_0^T \Big(\hat{G}^i(t) - g^*(\hat{X}^i(t))\Big)\,dt\Big|^2\right] = 0, \tag{4.5.4}$$

and this is not sufficient to conclude that $(\hat{X}^i)_{i\in\mathbb{N}}$ is tight. In fact, in Remark 9.5 from Chapter 3 of Ethier & Kurtz [16] an example is constructed of $\{0,1\}$-valued processes X_n with generators G_n, for which we can check that

$$\sup_{n=1,2,\ldots} E\left[\left|\int_0^T (G_n f)(X_n(t))dt\right|^2\right] < \infty, \tag{4.5.5}$$

for all $f \in \mathcal{C}(E)$, while the sequence $(X_n)_{n=1,2,\ldots}$ is not tight in $\mathcal{D}_{\{0,1\}}[0,\infty)$.

We next investigate why our arguments give us (4.5.4) and not (4.5.3). The essential step in the proof is Lemma 4.4.7. For $i \in \mathbb{N}$ let X^{N_i} be a solution of (4.1.5) with initial condition (4.4.22) and let let τ_i ($i \in \mathbb{N}$) be exponentially distributed random variables as in Section 4.4.2, with means λ_i satisfying (4.4.94). For $i \in \mathbb{N}$ and $\xi \in \Omega_{N_i}$ and $t \geq 0$ let us write

$$Z^i_\xi(t) := \sigma_{k_i} g(X_0^{N_i}(t)) - g^*(x_0^{k_i}). \tag{4.5.6}$$

It follows from Lemma 4.4.6 that there exist constants M_i, tending to zero as $t \to \infty$, such that for all $x \in K^{\Omega_i}$ and all $\omega \in \mathcal{D}_\mathbb{R}[0,\infty)$ with $\omega(0) = \sigma_{k_i} g(x_0) - g^*(x_0^{k_i})$:

$$E[Z^i_0(\tau_i) | (Z^i_{\xi_i}(t))_{t\geq 0}] \leq M_i. \tag{4.5.7}$$

As is shown in Lemma 4.4.7, this implies that there exist constants M_i, tending to zero as $t \to \infty$, such that for all $x \in K^{\Omega_{N_i}}$

$$E[Z^i_0(\tau_i) Z^i_{\xi_i}(\tilde{\tau}_i)] \leq M_i, \tag{4.5.8}$$

where τ_i, $\tilde{\tau}_i$ are two indepenent exponentially distributed random variables with mean λ_i. As is shown in Lemma 4.4.8, formula (4.5.8) leads to (4.5.4). For (4.5.3) we would need that

$$E[Z^i_0(\tau_i) Z^i_{\xi_i}(\tau_i)] \leq M_i, \tag{4.5.9}$$

where the same τ_i occurs twice. This conclusion cannot be drawn, however, from (4.5.7). To give a simple example of what may go wrong, consider random variables $Y(1), Y(2), Z(1), Z(2)$, taking values in $\{-1, 1\}$, and independent of these a random variable τ that with probabilities $\frac{1}{2}$ takes the values 1 and 2. Assume that

$$(Y(1), Y(2),\ Z(1), Z(2)) = \begin{cases} (\ \ 1, -1,\ -1, -1) & \text{with probability } \varepsilon \\ (\ \ 1, -1,\ \ \ 1, -1) & \text{with probability } \frac{1}{2} - \varepsilon \\ (-1,\ \ 1, -1,\ \ 1) & \text{with probability } \frac{1}{2} - \varepsilon \\ (-1,\ \ 1,\ \ 1,\ \ 1) & \text{with probability } \varepsilon. \end{cases} \tag{4.5.10}$$

Then

$$E[Y(\tau)|Z(1) = z_1,\ Z(2) = z_2] = 0 \qquad (z_1, z_2 \in \{-1, 1\}), \tag{4.5.11}$$

while

$$\lim_{\varepsilon \to 0} E[Y(\tau)Z(\tau)] = 1 \neq 0. \tag{4.5.12}$$

Appendix A

Diffusions on compact state space

A.1 Why this appendix?

There exist many texts that offer to the mathematical reader an introduction to the theory of stochastic differential equations. Some of the books that were advised to me and that I used are the ones by Breiman [3], Chung & Williams [4], Dynkin [15], Ethier & Kurtz [16], Hida [20], Karatzas & Shreve [22], Oeksendal [27], Revuz & Yor [31] and Stroock & Varadhan [40]. In spite of all the available material, I never managed to find a text that would give a trained functional analyst or probabalist in an afternoon's time a rough idea of what I have come to consider as the basic ideas of the theory; what stochastic differential equations were invented for, and what they can do. This is probably due partly to my personal preferences and partly to the complex history of the subject.

The study of diffusion processes started in the 1930's, when Kolmogorov and Feller used methods from the theory of partial differential equations to establish the existence and uniqueness of probability densities of diffusion processes. Their functional analytic approach led to a complete characterization of all Feller processes in one dimension, but for higher-dimensional domains there remained major problems. Following another line of research, Itô started in the 1940's his pioneering work on stochastic differential equations. It was not until Yamada and Watanabe proved in 1971 that strong uniqueness implies weak uniqueness, that finally a solid link between the two approaches was made. A few years earlier, Stroock and Varadhan had invented the martingale problem as a means to use results from the analytic approach to prove weak uniqueness of stochastic differential equations.

The historic split between the two approaches still influences today's literature. Thus, a book like Oeksendal's [27] offers a quick and easy-to-read introduction into Itô's theory of stochastic differential equations, but tells little about the functional analytic approach. As a result, certain concepts must remain somewhat mysterious to the reader: Why, for example, the two different concepts of uniqueness, and why is it the square of the diffusion matrix that really matters? Other books, like Karatzes & Shreve's [22], give the full story, but here the results linking the two approaches are presented as some further developments in the theory, where they may easily be overlooked by the beginning reader. The book that I at present consult most is the one by Ethier & Kurtz [16], which presents the material in the order that I consider to be most natural: first the analytic theory, then the martingale problem, and then stochastic differential equations. Apart from that, it also contains some strong theorems that are elsewhere absent. However, like many of its fellow books on the subject, it offers no easy reading, because of the technical nature of the material and the fact that theorems are stated in a highly general form. It has more than once taken me two handwritten pages to check that a proposition I wished to verify in a concrete setting, really was a straightforward consequence of a Theorem, a Remark and a Problem in the book.

This appendix tries to take the functional analyst or probabilist with no experience in the field of diffusion processes by the hand and show him or her around in the shortest possible time. I hope that those who are more familiar with the material will find it pleasant reading too. In view of the material covered in this disseration, we take one specific problem as our motivation: how to find out if there exists a unique Feller diffusion on a compact and convex domain with a given drift and diffusion function.

By restricting ourselves to compact domains, we can avoid a lot of the technicalities that clutter the general theorems in books. The restriction to convex domains is not very essential, but facilitates the discussion of boundary behavior. The type of processes we consider do not need boundary conditions to be specified in terms of a restriction on the domain of the generator; instead, the drift and the diffusion function suffice to specify the behavior of the process, not only in the interior of the domain but also on the boundary. This type of processes is still fairly ill understood. Thus, we will see that the theory of diffusion processes after more than half a century still holds some fundamental unsolved problems.

Most of the propositions below can be found in, or easily deduced from, Ethier & Kurtz' book [16], to which we refer for proofs and technical details.

A.2 Transition probabilities

Let E be a compact separable metrizable space. We denote the Borel-σ-field (the σ-field generated by the open sets) by $\mathcal{B}(E)$, and we write $\mathcal{P}(E)$ for the space of all probability measures on $\mathcal{B}(E)$, equipped with the topology of weak convergence.

A continuous transition probability on E is a function $P_t(x, dy)$ with the following properties:

1. $(t, x) \mapsto P_t(x, \cdot)$ is a continuous map from $[0, \infty) \times E$ into $\mathcal{P}(E)$
2. $P_0(x, \cdot) = \delta_x$ (the delta-measure in x)
3. $\displaystyle\int_E P_t(x, dy) P_s(y, dz) = P_{t+s}(x, dz)$.

The third property is called the *Chapman-Kolmogorov property*. We think of $P_t(x, dy)$ as describing the probability that a certain continuous-time Markov process takes values in the interval dy, given that it was a time t before at the position x. Indeed, we have the following:

Proposition A.2.1 *Let $P_t(x, dy)$ be a continuous transition probability, and let $\mu \in \mathcal{P}(E)$. Then there exists a Markov process $(X(t))_{t \geq 0}$, unique in distribution, such that*

1. $P[X(0) \in dx] = \mu(dx)$
2. $P[X(s+t) \in dy | X(s) = x] = P_t(x, dy)$.

Markov processes with such continuous transition probabilities are called *Feller processes* (with state space E), and we would like to know how to construct them. Because of the Chapman-Kolmogorov property, we do not need to know the transition probabilities $P_t(x, dy)$ for all t in order to specify the process uniquely. In fact, we have the feeling that it should somehow be enough to know $P_t(x, dy)$ for infinitesimal t.

For example, we can look at a process X for which

$$P_t(x, dy) = \delta_x(dy) + \lambda\Big(\Gamma(x, dy) - \delta_x(dy)\Big)t + o(t), \tag{A.2.1}$$

where λ is a positive number and $\Gamma(x, dy)$ is a continuous probability kernel on E, which means that $x \mapsto \Gamma(x, \cdot)$ is a continuous map from E into $\mathcal{P}(E)$. The precise meaning of the small o-notation will become clear in the next section. A Feller process with transition probabilities as in (A.2.1) is called a *jump process*. It is a process that stays at a position x during an exponential time with mean λ^{-1}, and then chooses a new place to jump to, according to the probability $\Gamma(x, \cdot)$. It is customary to choose a version of X such that it has piecewise constant right-continuous sample paths.

One can also look at processes that have continuous paths. Consider the case that E is a compact and convex subset of $\mathbb{R}^d$. A *diffusion process* on E is a Feller process X whose transition probabilities satisfy

$$\begin{aligned} &1. \quad \int_E P_t(x,dy)(y_i - x_i) = b_i(x)\,t + o(t) \\ &2. \quad \int_E P_t(x,dy)(y_i - x_i)(y_j - x_j) = a_{ij}(x)\,t + o(t) \\ &3. \quad \int_E P_t(x,dy)1_{\{|x-y|>\varepsilon\}} = o(t), \end{aligned} \tag{A.2.2}$$

uniformly in x as $t \to 0$ for all $i, j = 1, \ldots, d$ and $\varepsilon > 0$. Here $b : E \to \mathbb{R}^d$ and $a : E \to \mathbb{R}^d \otimes \mathbb{R}^d$ are continuous functions taking values in the d-dimensional real vectors and the $d \times d$ positive symmetric real matrices, respectively. The function b is called the *drift function* (or 'local drift function' or simply 'drift'). $\sum_i b_i(x)z_i$ measures the tendency of the process at position x to move in a direction z. The function a is called the *diffusion function* (or 'local diffusion function' or 'diffusion matrix'). $\sum_{ij} z_i a_{ij}(x) z_j$ measures the strength of the random fluctuations of the process at position x in the direction z.

Property 3 in formula A.2.2 expresses the continuity of the process. It can be shown (although this is not easy to see at this stage) that a Feller process X satisfying the properties 1 and 2 in (A.2.2) has continuous sample paths if and only if it satisfies Property 3.

The idea behind (A.2.2) is that in order to specify a probability distribution that is very strongly peaked around x, it should be sufficient to know only its first and second order moments. We expect that if we are given the first and second order moments of small increments $X(s+t) - X(s)$ of our process, then these will by some sort of Central Limit Theorem uniquely determine the distribution of a large sum of such increments, and hence of our whole process.

This intuition is largely correct: in many cases it is possible to prove that to a given drift function and diffusion function there corresponds a unique (in distribution) diffusion process. The methods needed to prove this are, however, far from trivial, as we will see in the remainder of this appendix.

In order for a process satisfying (A.2.2) to exist, we have to put restrictions on b and a that guarantee that the process, so to say, does not want to leave the space E. For any $x \in E$ we denote by N_x the cone of normal vectors to E in x:

$$z \in N_x \quad \Leftrightarrow \quad \sum_i (y_i - x_i) z_i \leq 0 \;\; \forall y \in E. \tag{A.2.3}$$

We denote by I_x the space of directions in which E is flat at x:

$$z \in I_x \quad \Leftrightarrow \quad \exists \varepsilon > 0 \text{ such that } x + \lambda z \in E \;\forall |\lambda| \leq \varepsilon\}. \tag{A.2.4}$$

Sufficient conditions on b and a turn out to be[1]

$$\begin{aligned} \sum_i b_i(x) z_i &\le 0 & \forall z \in N_x \\ \sum_{ij} z_i a_{ij}(x) z_j &= 0 & \forall z \in I_x^\perp. \end{aligned} \tag{A.2.5}$$

Here $I_x^\perp$ is the orthogonal complement of I_x. Note that not every vector in $I_x^\perp$ is in the linear span of N_x. (For example, if E is the unit ball and $|x| = 1$, then $I_x^\perp = \mathbb{R}^d$, while N_x is one-dimensional.) Thus the second condition in (A.2.5) is stronger than the requirement that the fluctuations in directions normal to the surface are zero.

A.3 Feller semigroups

We write $\mathcal{C}(E)$ for the Banach space of continuous real functions on E, equipped with the supremum norm. Whenever $P_t(x, dy)$ is a continuous transition probability, the formula

$$(S_t f)(x) := \int_E P_t(x, dy) f(y) \qquad (x \in E,\ f \in \mathcal{C}(E)) \tag{A.3.1}$$

defines a *Feller semigroup* on $\mathcal{C}(E)$. Here, a family of operators $(S_t)_{t \ge 0}$ on $\mathcal{C}(E)$ is called a Feller semigroup if

1. For each $t \ge 0$, $S_t : \mathcal{C}(E) \to \mathcal{C}(E)$ is a linear operator
2. $S_t 1 = 1$ for all $t \ge 0$
3. $f \ge 0 \Rightarrow S_t f \ge 0$ for all $t \ge 0$
4. $\lim_{t \to 0} \|S_t f - f\| = 0$
5. $S_t S_s = S_{t+s}$ for all $t, s \ge 0$ and S_0 is the identity.

Conversely, one can see that to each Feller semigroup there corresponds a unique continuous transition probability. By definition, the *generator* of a Feller semigroup is the operator

$$(Gf)(x) := \lim_{t \to 0} t^{-1}(S_t f - f), \tag{A.3.2}$$

[1]What one really needs is that the operator A in (A.3.5) below satisfies the maximum principle. For this, condition (A.2.5) is sufficient, but not necessary. When formulated as separate conditions on b and a, (A.2.5) is as far as one can go, but for certain combinations of b and a less may suffice. For example, if $E = \{x \in \mathbb{R}^2 : |x| \le 1\}$, $b_i(x) = -cx_i$ and $a_{11}(x) = x_2 x_2$, $a_{12}(x) = a_{21}(x) = -x_1 x_2$, $a_{22}(x) = x_1 x_1$, then the operator A in (A.3.5) satisfies the maximum principle if and only if $c \ge \frac{1}{2}$.

with domain $\mathcal{D}(G)$ the space of all functions $f \in \mathcal{C}(E)$ for which the limit in (A.3.2) exists in the topology on $\mathcal{C}(E)$. It is a theorem that Feller semigroups are uniquely determined by their generator. In that sense our earlier intuition is justified that in order to specify a continuous transition probability $P_t(x, dy)$, it is enough to give it for infinitesimal t. We now need a criterion to see whether an operator G is the generator of a Feller semigroup.

We say that an operator A on $\mathcal{C}(E)$ with domain $\mathcal{D}(A)$ satisfies the maximum principle if, whenever a function $f \in \mathcal{D}(A)$ assumes its maximum over E in a point $x \in E$, we have $(Af)(x) \leq 0$. We say that A is closed if and only if its graph $\{(f, Af) : f \in \mathcal{D}(A)\}$ is a closed subset of $\mathcal{C}(E) \times \mathcal{C}(E)$. The following is a version of the Hille-Yosida theorem:

Proposition A.3.1 *A linear operator G on $\mathcal{C}(E)$ is the generator of a Feller semigroup if and only if*

1. $1 \in \mathcal{D}(G)$ *and* $G1 = 0$
2. G *satisfies the maximum principle*
3. $\mathcal{D}(G)$ *is dense in* $\mathcal{C}(E)$
4. *For every $f \in \mathcal{D}(G)$ there exists a continuously differentiable function $t \mapsto f_t$ such that $f_0 = f$, $f_t \in \mathcal{D}(G)$ and $\frac{\partial}{\partial t} f_t = G f_t$*
5. G *is closed.*

Here the differentiation with respect to t is in the Banach space $\mathcal{C}(E)$. The difficult condition in the Hille-Yosida theorem is condition 4. We need to show that the differential equation $\frac{\partial}{\partial t} f_t = G f_t$ with initial condition $f_0 = f$ has a solution. It then follows from the general theory that this solution is unique, and equals $f_t = S_t f$.

For jump processes as in (A.2.1) we can do this. We define an operator G with domain $\mathcal{D}(G) = \mathcal{C}(E)$ by

$$(Gf)(x) := \lambda\Big(\int_E \Gamma(x, dy) f(y) - f(x) \Big). \tag{A.3.3}$$

This is a bounded linear operator on $\mathcal{C}(E)$, and for each $t \geq 0$ the infinite sum

$$e^{tG} := \sum_{n=0}^{\infty} \frac{1}{n!} (tG)^n \tag{A.3.4}$$

converges. It is now easy to see that $f_t := e^{tG} f$ solves the equation $\frac{\partial}{\partial t} f_t = G f_t$, and therefore G is the generator of a Feller semigroup, which in this simple case is just given by $S_t = e^{tG}$.

For diffusion processes, life is a lot harder. We cannot expect the domain of their generator to be all of $\mathcal{C}(E)$. In fact, it is not at all clear what $\mathcal{D}(G)$ is. Let A be the operator

$$(Af)(x) := \sum_i b_i(x)\tfrac{\partial}{\partial x_i} f(x) + \tfrac{1}{2}\sum_{ij} a_{ij}(x)\tfrac{\partial^2}{\partial x_i \partial x_j} f(x), \tag{A.3.5}$$

with domain

$$\mathcal{D}(A) := \mathcal{C}^2(E), \tag{A.3.6}$$

the space of real functions on E that can be extended to functions in $\mathcal{C}^2(\mathbb{R}^d)$. Properties 1 and 2 in (A.2.2) imply that G should be an extension of the operator A_0, which is the restriction of A to the smaller domain

$$\mathcal{D}(A_0) = \text{the polynomials of degree} \leq 2. \tag{A.3.7}$$

Using Property 3 in (A.2.2) and a Taylor expansion it is possible to show (see section B.1) that $Gf = Af$ for every $f \in \mathcal{C}^2(E)$, so G has to be an extension of the operator A.

We can easily see that the operator A itself is not the full generator G of a Feller semigroup. In fact, A is not closed. But we can hope that maybe its closure is the generator of a Feller semigroup. Here, the following version of the Hille-Yosida theorem comes to our help.

Proposition A.3.2 *Assume that a linear operator A on $\mathcal{C}(E)$ satisfies*

1. $1 \in \mathcal{D}(A)$ *and* $A1 = 0$
2. *A satisfies the maximum principle*
3. *$\mathcal{D}(A)$ is dense in $\mathcal{C}(E)$*
4. *There exists a dense subspace $D \subset \mathcal{C}(E)$ with the property that for every $f \in D$ there exists a continuously differentiable function $t \mapsto f_t$ such that $f_0 = f$, $f_t \in \mathcal{D}(A)$ and $\frac{\partial}{\partial t} f_t = Af_t$.*

Then the closure of A generates a Feller semigroup.

For our operator A, the first three conditions in this proposition are easily checked, where the fact that A satisfies the maximum principle follows from the containment condition (A.2.5). In order to check condition 4, we must find solutions of a *Cauchy equation*:

$$\tfrac{\partial}{\partial t} f_t(x) - \sum_i b_i(x)\tfrac{\partial}{\partial x_i} f_t(x) - \tfrac{1}{2}\sum_{ij} a_{ij}(x)\tfrac{\partial^2}{\partial x_i \partial x_j} f_t(x) = 0, \tag{A.3.8}$$

with initial condition $f_0(x) = f(x)$. We have to find solutions of (A.3.8) for all f in some dense space D. This is in general very hard.

We can make life a little easier by applying a Laplace transform with respect to our time variable. It is in this language that the Hille-Yosida theorem is usually stated.

Proposition A.3.3 *The closure $\overline{A}$ of a linear operator A on $\mathcal{C}(E)$ is the generator of a Feller semigroup if and only if*

1. $1 \in \mathcal{D}(\overline{A})$ *and* $\overline{A}1 = 0$
2. *A satisfies the maximum principle*
3. *$\mathcal{D}(A)$ is dense in $\mathcal{C}(E)$*
4. *There exists a $\lambda \in (0, \infty)$ and a dense subspace $D \subset \mathcal{C}(E)$ with the property that for every $f \in D$ there exists a $p_\lambda \in \mathcal{D}(A)$ such that $(1 - \lambda A)p_\lambda = f$.*

Thus, the closure of the operator A in (A.3.5) generates a Feller semigroup if and only if we can find solutions of a *Laplace equation*:

$$p_\lambda(x) - \lambda \sum_i b_i(x)\tfrac{\partial}{\partial x_i} p_\lambda(x) - \lambda \tfrac{1}{2} \sum_{ij} a_{ij}(x) \tfrac{\partial^2}{\partial x_i \partial x_j} p_\lambda(x) = f(x), \tag{A.3.9}$$

for a dense collection of functions f. It is not hard to show that if f_t is a solution to the Cauchy equation (A.3.8), then

$$p_\lambda := \int_0^\infty f_t \lambda^{-1} e^{-\frac{t}{\lambda}} dt \tag{A.3.10}$$

is a solution to the Laplace equation (A.3.9). However, having a solution to (A.3.9) for one particular value of λ (as is required in the Hille-Yosida theorem) does not automatically imply that we can solve (A.3.8).

Although the Laplace-version of the Hille-Yosida theorem is some improvement over the Cauchy-version, it does not make much of a difference in practice. For the type of diffusion processes that we are considering in this appendix, there are almost no results showing existence of solutions to (A.3.9).

There are results, however, for Cauchy and Laplace problems in a setting that is slightly different from ours. These results cover the case where for all x the positive symmetric matrix $a_{ij}(x)$ is strictly positive. In this case, the conditions (A.2.5) are violated, and one has to find other ways of preventing the process from leaving E. This can be done by restricting the domain of the operator A by certain boundary conditions. These boundary conditions guarantee that A, on its restricted domain, satisfies the maximum principle. It can be seen that different types of boundary conditions correspond to different types of boundary behavior, such as absorbing boundaries, reflecting boundaries, and so on. For such operators A, existence of solutions to the Cauchy problem has been shown, at least for initial conditions $f \in \mathcal{C}^2(E)$ and for drift and diffusion functions that are Hölder continuous.

To give an example where in our setting (A.3.8) has a solution for a dense collection of initial conditions, consider the case where the operator A has the form

$$(Af)(x) = -\sum_i x_i \tfrac{\partial}{\partial x_i} f(x) + (1 - |x|^2) \sum_i \tfrac{\partial^2}{\partial x_i{}^2} f(x), \tag{A.3.11}$$

It is easy to see that A maps a polynomial of degree $\leq n$ to a polynomial of degree $\leq n$. Using this fact one can show that the Cauchy problem has a polynomial solution for each polynomial initial condition. It follows that the closure of A generates a Feller semigroup on $\mathcal{C}(E)$, where the corresponding Feller process can be shown to have continuous sample paths. To do something more general we need different techniques.

A.4 The martingale problem

In the course of the last section, we changed the question that we originally wanted to answer into a more specific question, that we subsequently were unable to solve. What we *really* wanted to know is: "Does there exist a unique Feller process with continuous sample paths whose generator is an extension of the operator A_0 in (A.3.7)?". The question that we finally adressed was: "Does the closure of the operator A in (A.3.5) generate a Feller semigroup?". This is a different question for two reasons. First, the operator A is not the same as the operator A_0. Second, even if there exists a unique Feller process with continuous sample paths whose generator extends a certain operator, then there is no reason why this generator should be the *closure* of the operator. This is already clear from the fact that the operator A_0 is closed but does not have a dense domain, so that its closure will never be the generator of a Feller semigroup. In this section we look for methods that do not require our generator to be the closure of some given operator.

For $\mu \in \mathcal{P}(E)$ and $f \in \mathcal{C}(E)$, let us introduce the 'dual' notation

$$\langle \mu | f \rangle := \int_E \mu(dx) f(x). \tag{A.4.1}$$

If G is the generator of a Feller semigroup and $P_t^x(dy) = P_t(x, dy)$ is the associated transition probability, then P_t^x is the unique solution of the equation

$$\tfrac{\partial}{\partial t} \langle P_t^x | f \rangle = \langle P_t^x | G f \rangle \qquad \forall f \in \mathcal{D}(G) \tag{A.4.2}$$

with initial condition $P_0^x = \delta_x$. We can think of (A.4.2) as the equation

$$\tfrac{\partial}{\partial t} P_t^x = G^* P_t^x, \tag{A.4.3}$$

where the adjoint operator $G^* : \mathcal{C}(E)^* \to \mathcal{C}(E)^*$ is defined by

$$\langle G^*\mu | f\rangle := \langle \mu | Gf\rangle \qquad (f \in \mathcal{D}(G)). \tag{A.4.4}$$

For a general linear operator A on $\mathcal{C}(E)$, we may wonder if the equation

$$\begin{array}{ll} \tfrac{\partial}{\partial t}\langle P_t^x | f\rangle = \langle P_t^x | Af\rangle & \forall f \in \mathcal{D}(A) \\ \langle P_0^x | f\rangle = f(x) & \forall f \in \mathcal{C}(E) \end{array} \tag{A.4.5}$$

has a unique solution for each $x \in E$. If A is the operator in (A.3.5), then (A.4.5) is the dual equation to the Cauchy equation (A.3.8).

There exist interesting relations between equations and their dual equations. For example, existence of solutions to (A.4.5) implies uniqueness of solutions to (A.3.8), while existence of solutions to (A.3.8) implies uniqueness of solutions to (A.4.5). To prove the latter, all one needs to do is to show that

$$\tfrac{\partial}{\partial t}\langle P_t^x | f_{T-t}\rangle = \langle P_t^x | A f_{T-t}\rangle - \langle P_t^x | A f_{T-t}\rangle = 0 \tag{A.4.6}$$

whenever P_t^x solves (A.4.5) and f_t solves (A.3.8). It then follows that

$$\langle P_T^x | f\rangle = \langle P_T^x | f_{T-T}\rangle = \langle P_0^x | f_T\rangle = f_T(x) \quad \forall f \in D, \tag{A.4.7}$$

and since $D \subset \mathcal{C}(E)$ is dense this relation determines P_T^x uniquely. However, existence of solutions to (A.3.8) is by no means necessary for the uniqueness of solutions to (A.4.5), and it is this fact that we will exploit. The idea will be, more or less, to show that whenever solutions to (A.4.5) exist and are unique, they are continuous transition probabilities. Then these transition probabilities give us a Feller semigroup, whose generator G is an extension of A, while it need not be the case that G is the closure of A.

In fact, we are not going to do precisely that, but something similar. Any Feller process $(X(t))_{t\geq 0}$ can be chosen such that it has sample paths that are continuous from the right and have left limits: the so-called cadlag functions. We denote the space of E-valued cadlag functions by $\mathcal{D}_E[0,\infty)$ and we equip it with the Skorohod topology and the associated Borel-σ-field. We can view the whole process $X = (X(t))_{t\geq 0}$ as a stochastic variable taking values in $\mathcal{D}_E[0,\infty)$ (it turns out that X is measurable with respect to $\mathcal{B}(\mathcal{D}_E[0,\infty))$).

Let A be a linear operator on $\mathcal{C}(E)$ with domain $\mathcal{D}(A)$. We say that a process X taking values in $\mathcal{D}_E[0,\infty)$ solves the *martingale problem* for A if for every $f \in \mathcal{D}(A)$ the process

$$M(t) := f(X(t)) - \int_0^t (Af)(X(s))ds \tag{A.4.8}$$

is a martingale with respect to the filtration generated by X. Note that $E[M(t)] = E[M(0)]$, so that (A.4.8) implies

$$E[f(X_t)] - E[f(X(0))] = \int_0^t E[(Af)(X(s))]ds \qquad \forall f \in \mathcal{D}(A). \tag{A.4.9}$$

Differentiating with respect to t, we see that the marginals

$$P_t^x(dy) := P[X^x(t) \in dy] \tag{A.4.10}$$

solve equation (A.4.5) whenever X^x solves the martingale problem for A with initial condition $X^x(0) = x$. Thus, a solution to the martingale problem for A is something like a solution to equation (A.4.5), endowed with a richer structure. The fact that P_t^x solves (A.4.5) is implied by the following more general result.

Proposition A.4.1 *Assume that G is the generator of a Feller semigroup and that $\mu \in \mathcal{P}(E)$. Let X be the unique (in distribution) Feller process with generator G and initial condition μ. Then X is the unique (in distribution) $\mathcal{D}_E[0,\infty)$-valued process with initial condition μ that solves the martingale problem for G.*

We say that *uniqueness* holds for the martingale problem for an operator A if for every $\mu \in \mathcal{P}(E)$ there exists at most one (in distribution) solution to the martingale problem for A with initial condition μ. We say that *existence* holds for the martingale problem for A if for every probability measure $\mu \in \mathcal{P}(E)$ there exists at least one solution to the martingale problem for A with initial condition μ. When both uniqueness and existence hold, we say that the martingale problem for A is *well-posed*.

Our next aim is to show that if the martingale problem for A is well-posed, then A has a unique extension to a generator of a Feller semigroup. If the closure of A generates a Feller semigroup, then it turns out that the martingale problem for A is well-posed, but the converse is in general not true.

The techniques that we need concern compactness and convergence of solutions to martingale problems. Let $\mathcal{P}(\mathcal{D}_E[0,\infty))$ denote the space of probability measures on $\mathcal{D}_E[0,\infty)$, equipped with the (metrizable) topology of weak convergence (defined with bounded continuous functions on $\mathcal{D}_E[0,\infty)$). We say that a sequence $(X_n)_{n=1,2,\ldots}$ of processes is *tight* if the collection of their laws is relatively compact in $\mathcal{P}(\mathcal{D}_E[0,\infty))$. Clearly, every tight sequence has a weakly convergent subsequence. We denote weak convergence of processes as well as other random variables by $\Rightarrow$.

Let $(A_n)_{n=1,2,\ldots}$ be a sequence of linear operators on $\mathcal{C}(E)$. We define the *extended limit* of $(A_n)_{n=1,2,\ldots}$ as

$$\underset{n\to\infty}{\text{ex-lim}}\, A_n := \{(f,g) : \exists f_n \in \mathcal{D}(A_n) \text{ such that } f_n \to f,\ Af_n \to g\}. \tag{A.4.11}$$

As a subspace of $\mathcal{C}(E)\times\mathcal{C}(E)$, $A := \text{ex-lim}_{n\to\infty} A_n$ always exists. Such a subspace A is sometimes called a multi-valued linear operator on $\mathcal{C}(E)$. If it has the property that $(f, g) \in A$, $(f', g) \in A \Rightarrow f = f'$, then A is 'single-valued' and we can associate it with a linear operator in the usual sense. For multi-valued operators A and B we say that $A \subset B$ if A is a subspace of B.

Proposition A.4.2 *Assume that A_n ($n = 1, 2, \ldots$) and A are linear operators on $\mathcal{C}(E)$ and that X_n are processes solving the martingale problem for A_n. Assume that $\mathcal{D}(A)$ is dense in $\mathcal{C}(E)$ and that*

$$\text{ex-}\lim_{n\to\infty} A_n \supset A. \tag{A.4.12}$$

Then the following hold:

1. *$(X_n)_{n=1,2,\ldots}$ is tight.*
2. *If $X_n \Rightarrow X$ for some X, then also $X_n(0) \Rightarrow X(0)$, and X solves the martingale problem for A.*

This proposition has many applications. First of all, it can be used to show that if the martingale problem for A is well-posed, then A has a unique extension to a generator of a Feller semigroup. For this, it suffices to show that

$$P_t(x, dy) := P[X^x(t) \in dy] \tag{A.4.13}$$

is a continuous transition probability when X^x is the solution of the martingale problem with initial condition $X^x(0) = x$. To see that $x \mapsto P_t(x, dy)$ is continuous, pick some $x_n \to x$. The X^{x_n} are tight by Proposition A.4.2, and therefore there exists a subsequence $x_{n(k)}$ such that $X^{x_{n(k)}} \Rightarrow X$ for some solution X to the martingale problem for A. Moreover, $X^{x_{n(k)}}(0) \Rightarrow X(0)$ so X has initial condition $X(0) = x$. By the uniqueness of solutions to the martingale problem, it therefore follows that $X = X^x$. Since the same is true for any convergent subsequence, we conclude that $X^{x_n} \Rightarrow X^x$. We can now also prove that $X^{x_n}(t) \Rightarrow X^x(t)$ for each $t \geq 0$, and therefore $P_t(x_n, \cdot) \Rightarrow P_t(x, \cdot)$. The continuity of $P_t(x, dy)$ in x and t together is shown in a similar way. It is trivial to see that $P_0(x, \cdot) = \delta_x$ and, finally, the Chapman-Kolmogorov property can be deduced too. It thus follows that:

Proposition A.4.3 *Let A be a linear operator on $\mathcal{C}(E)$. Assume that $\mathcal{D}(A)$ is dense in $\mathcal{C}(E)$ and that the martingale problem for A is well-posed. Then there exists a unique Feller semigroup with the property that its generator G is an extension of A. This semigroup is given by*

$$(S_t f)(x) = E[f(X^x(t))], \tag{A.4.14}$$

where X^x is the solution of the martingale problem for A with initial condition $X^x(0) = x$.

Proposition A.4.2 may also be used to show existence of solutions to the martingale problem for a "difficult" operator A, by approximating A with "easy" operators. For example, if we can find generators G_n ($n = 1, 2, \ldots$) of jump processes as in (A.2.1) such that

$$\text{ex-}\lim_{n\to\infty} G_n \supset A, \tag{A.4.15}$$

then for each $n = 1, 2, \ldots$ there exist solutions X_n to the martingale problem for G_n, and by Proposition A.4.2 the sequence $(X_n)_{n=1,2,\ldots}$ has accumulation points, where each accumulation point X solves the martingale problem for A. Using much heavier techniques (involving measurable semigroups that are not necessarily Feller), but based on the same approximation idea, the following result is derived:

Proposition A.4.4 *Assume that a linear operator A on $\mathcal{C}(E)$ satisfies*

1. $1 \in \mathcal{D}(A)$ *and* $A1 = 0$
2. *A satisfies the maximum principle*
3. *$\mathcal{D}(A)$ is dense in $\mathcal{C}(E)$.*

Then existence holds for the martingale problem for A.

Note that these are just the three easy conditions in Proposition A.3.3. In particular, Proposition A.4.4 implies the existence of solutions to the martingale problem for the operator A in (A.3.5), whenever the drift an diffusion function are continuous and satisfy (A.2.5). We are left with the task to verify that they are unique.

A.5 Stochastic differential equations

Assume that $E \subset \mathbb{R}^d$ is compact and convex, and that $b : E \to \mathbb{R}^d$ and $\sigma : E \to \mathbb{R}^d \otimes \mathbb{R}^d$ are continuous functions taking values in the d-dimensional real vectors and the $d \times d$ real matrices, respectively. It is possible to give a precise mathematical meaning to the *stochastic differential equation*

$$dX_i(t) = b_i(X(t))\,dt + \sum_j \sigma_{ij}(X(t))\,dB_j(t) \qquad (t \geq 0,\ i = 1, \ldots, d). \tag{A.5.1}$$

Here $B = (B_1(t), \ldots, B_d(t))_{t\geq 0}$ is d-dimensional Brownian motion, adapted to a certain given filtration $(\mathcal{F}_t)_{t\geq 0}$, and the equation (A.5.1) should be read as

$$X_i(t) - X_i(0) = \int_0^t b_i(X(t))\,ds + \sum_j \int_0^t \sigma_{ij}(X(s))\,dB_j(s) \quad \text{a.s.}, \tag{A.5.2}$$

where the second integral is a so-called stochastic integral. By the compactness of E, the integrand $\sigma_{ij}(X(s))$ is bounded, and for any $\overline{\mathcal{F}}_t$-adapted process X with continuous sample paths, the right-hand side in (A.5.2) is an almost surely defined continuous $\overline{\mathcal{F}}_t$-adapted process. Here $\overline{\mathcal{F}}_t$ is the smallest σ-algebra containing $\mathcal{F}_t$ and all measurable null-sets.

We denote the space of continuous functions from $[0, \infty)$ to E by $\mathcal{C}_E[0, \infty)$. By a (weak) solution to (A.5.1) we mean a probability space $(\Omega, \mathcal{F}, P)$, with a filtration $(\mathcal{F}_t)_{t\geq 0}$ and an adapted d-dimensional Brownian motion $(B(t))_{t\geq 0}$, together with an $\overline{\mathcal{F}}_t$-adapted process X with sample paths in $\mathcal{C}_E[0, \infty)$, such that (A.5.2) holds.

A central result in the theory of stochastic integration is Itô's formula. In symbolic notation, it reads as follows.

Proposition A.5.1 *Assume that X and Y are processes with continuous sample paths satisfying*

$$dY_i(t) = b_i(X(t))\,dt + \sum_j \sigma_{ij}(X(t))\,dB_j(t). \tag{A.5.3}$$

If $f \in \mathcal{C}^2(E)$, then

$$df(Y(t)) = \sum_i (\tfrac{\partial}{\partial x_i} f)(Y(t))\,dY_i(t) + \tfrac{1}{2}\sum_{ij}(\tfrac{\partial^2}{\partial x_i \partial x_j} f)(Y(t))\,dY_i(t)\,dY_j(t), \tag{A.5.4}$$

where in evaluating $dY_i(t)\,dY_j(t)$ the following rules apply:

$$\begin{aligned} dt\,dt &= 0 \\ dt\,dB_i(t) &= 0 \\ dB_i(t)\,dB_j(t) &= \delta_{ij}dt. \end{aligned} \tag{A.5.5}$$

For us the following consequence is relevant: if X is a solution to the stochastic differential equation (A.5.1) and $f \in \mathcal{C}^2(E)$, then

$$\begin{aligned} f(X(t)) &- \sum_i \int_0^t b_i(X(s))(\tfrac{\partial}{\partial x_i} f)(X(s))\,ds \\ &-\tfrac{1}{2}\sum_{ij}\int_0^t \Big(\sum_k \sigma_{ik}(X(s))\sigma_{jk}(X(s))\Big)(\tfrac{\partial^2}{\partial x_j \partial x_j} f)(X(s))\,ds \\ &= f(X(0)) + \sum_{ij}\int_0^t \sigma_{ij}(X(s))(\tfrac{\partial}{\partial x_i} f)(X(s))\,dB_j(s). \end{aligned} \tag{A.5.6}$$

Here the stochastic integral yields a martingale, and thus we see that a solution X to the stochastic differential equation (A.5.1) solves the martingale problem for an operator A of the form (A.3.5), with

$$a_{ij}(x) = \sum_k \sigma_{ik}(x)\sigma_{jk}(x). \tag{A.5.7}$$

This result has a deep converse: it turns out that any solution to the martingale problem for A can be represented, on an appropriate space equipped with an appropriate Brownian motion, as a solution to the stochastic differential equation (A.5.1).

Proposition A.5.2 *Assume that X has sample paths in $\mathcal{C}_E[0,\infty)$ and solves the martingale problem for the operator A_0 in (A.3.7). If $\sigma : E \to \mathbb{R}^d \otimes \mathbb{R}^d$ is continuous and satisfies*

$$a_{ij}(x) = \sum_k \sigma_{ik}(x)\sigma_{jk}(x), \tag{A.5.8}$$

then there exists a solution $\tilde{X}$ to the stochastic differential equation (A.5.1) such that X and $\tilde{X}$ are equal in distribution.

We note that for a given continuous a it is always possible to find a continuous 'root' σ, that is, a σ that satisfies (A.5.8) (see section B.2).

The condition in Proposition A.5.2 that X has continuous sample paths can be dropped under mild conditions.

Proposition A.5.3 *Let A be a linear operator on $\mathcal{C}(E)$. Assume that for each $x \in E$ there exists an $f_x \in \mathcal{D}(A)$ such that for every $\varepsilon > 0$*

$$\inf_{\substack{x,y\in E \\ |x-y|>\varepsilon}} f_x(y) - f_x(x), \tag{A.5.9}$$

and for each $x \in E$

$$\lim_{y\to x} Af_y(x) = Af_x(x) = 0. \tag{A.5.10}$$

Then every solution to the martingale problem for A has sample paths in $\mathcal{C}_E[0,\infty)$.

In our case, the following choice of $f_x(y)$ works:

$$f_x(y) := |x-y|^{2+\varepsilon} \qquad (\varepsilon > 0). \tag{A.5.11}$$

Note, however, that the domain of A_0 is just a bit too small to contain functions f_x satisfying (A.5.9). For A there is of course no problem.

Now that the link between solutions of stochastic differential equations and solutions to the martingale problem is made, we investigate how this link can help us to answer the question that we were left with at the end of section A.4: When are solutions to the martingale problem for A unique? The following generalization of Proposition A.5.2 contains a central contribution of the theory of stochastic integration to the theory of Feller semigroups.

Proposition A.5.4 *Assume that X^1 and X^2 are solutions to the martingale problem for A_0 (defined on different probability spaces) with sample paths in $\mathcal{C}_E[0,\infty)$ and initial conditions $\mathcal{L}(X^1(0)) = \mathcal{L}(X^2(0))$, and assume that σ is a continuous root of a. Then there exist solutions $\tilde{X}^1$ and $\tilde{X}^2$ to the stochastic differential equation (A.5.1), defined on the same probability space and adapted to the same Brownian motion, such that $\tilde{X}^1(0) = \tilde{X}^2(0)$ a.s. and such that $\tilde{X}^\alpha$ is equal in distribution to X^α for $\alpha = 1, 2$.*

We say that *weak uniqueness* holds for the stochastic differential equation (A.5.1) if any two solutions X and Y with equal initial conditions (in law), defined on different probability spaces and adapted to different Brownian motions, have the same law. We say that *strong uniqueness* holds for the stochastic differential equation (A.5.1) if any two solutions X and Y with equal initial conditions (almost surely), defined on the same probability space and adapted to the same Brownian motion, are equal (almost surely).

By Propositions A.5.2 and A.5.3, weak uniqueness for the stochastic differential equation (A.5.1) is equivalent to uniqueness for the martingale problem for the operator A. Proposition A.5.4 shows that *strong uniqueness implies weak uniqueness*. The essence of Proposition A.5.4 is that it provides us with a *coupling technique* for showing uniqueness of solutions to the martingale problem for A. Using Itô's formula one sees that, with $\tilde{X}^1$ and $\tilde{X}^2$ as in Proposition A.5.4, the joint process $(\tilde{X}^1, \tilde{X}^2) = (\tilde{X}_i^\alpha)_{i=1,\ldots,d}^{\alpha=1,2}$ solves the martingale problem for the operator

$$(\hat{A}f)(x) := \sum_{i,\alpha} b_i(x^\alpha)\frac{\partial}{\partial x_i^\alpha} f(x) + \frac{1}{2}\sum_{ij,\alpha\beta}\Big(\sum_k \sigma_{ik}(x^\alpha)\sigma_{jk}(x^\beta)\Big)\frac{\partial^2}{\partial x_i^\alpha \partial x_j^\beta} f(x), \tag{A.5.12}$$

with domain $\mathcal{D}(\hat{A}) = \mathcal{C}^2(E \times E)$. Here we write $x = (x_i^\alpha)_{i=1,\ldots,d}^{\alpha=1,2}$ for a point in $E \times E$. Thus, the content of Proposition A.5.4 is that two solutions to the martingale problem for A can always be represented on the same space in such a way that the joint process solves the martingale problem for an operator $\hat{A}$ as in (A.5.12). Here we are still free in our choice of the continuous root σ of a.

We can now bring this knowledge into practice. A typical approach is the following. We can use Itô's formula (or, equivalently, the fact that $(\tilde{X}^1, \tilde{X}^2)$ solves the martingale problem for $\hat{A}$) to derive an equation for the time evolution of $E[|\tilde{X}^1(t) - \tilde{X}^2(t)|^2]$. If the functions $x \mapsto b(x)$ and $x \mapsto \sigma(x)$ are Lipschitz continuous with Lipschitz constant L, then a simple application of Gronwall's lemma gives the estimate

$$E[|\tilde{X}^1(t) - \tilde{X}^2(t)|^2] \leq E[|\tilde{X}^1(0) - \tilde{X}^2(0)|^2]\, e^{Lt}. \tag{A.5.13}$$

In particular, if $X^1(0) = X^2(0)$ then $X^1(t) = X^2(t)$ a.s. for all $t \geq 0$, and it follows that the martingale problem for A is well-posed.

If the space E is one-dimensional, then one can use the ordering of $\mathbb{R}$ to derive even stronger results. For strong uniqueness it now suffices to show that $\tilde{X}^1(0) \leq \tilde{X}^2(0)$ implies $\tilde{X}^1(t) \leq \tilde{X}^2(t)$ for all $t \geq 0$. We list two standard results on strong uniqueness, one for $d = 1$ and one for higher-dimensional state space.

Proposition A.5.5 *Let σ be a continuous root of a and assume that the function $x \mapsto b(x)$ is Lipschitz continuous. Then the following holds:*

1. *If $d = 1$ and the function $x \mapsto \sigma(x)$ is Hölder-$\frac{1}{2}$-continuous, then strong uniqueness holds for (A.5.1).*

2. *If $d \geq 2$ and the function $x \mapsto \sigma(x)$ is Lipschitz continuous, then strong uniqueness holds for (A.5.1).*

In this short overview we cannot do justice to the many different techniques that have been invented for showing uniqueness, weak or strong, to solutions of stochastic differential equations. The fact remains, however, that the results known in dimensions $d \geq 2$ are far from complete. There is a considerable gap between the class of drift and diffusion functions for which uniqueness is known to hold and those for which it is known to fail.

The difficulty with the scheme described above is that, even though the diffusion function a may be Lipschitz continuous, it may not be possible to choose a Lipschitz continuous root of a. As one may guess, problems arise when a is not strictly positive, as is the case on the boundary of E (see section B.2).

Let us for example return to the operator A introduced in (A.3.11). By the trick with the polynomials explained there, we know that the closure of A generates a Feller semigroup, and thus the martingale problem for A is well-posed. Can we derive this result here again, this time making use of Proposition A.5.5 and the deep results on the relation between the martingale problem and stochastic differential equations? In fact, we cannot. There simply exists no Lipschitz continuous root σ of a (Proposition B.2.2 below).

In this dissertation, we are led in a natural way to consider diffusion functions similar to the one in (A.3.11), for which the trick with the polynomials fails. For example, it may happen that A maps a polynomial of degree n into a polynomial of degree $n + 1$. For some specific initial conditions and choices of the drift function, we are able to prove uniqueness (see Theorem 2.2.9 and the Examples 3.1.8 and 3.1.9), but the general case remains open (see section 1.1.5).

Appendix B

Diffusion matrices

B.1 Convergence to second order partial differential operators

Proposition B.1.1 *Let E be a compact convex subset of $\mathbb{R}^d$ and let $b : E \to \mathbb{R}^d$ and $a : E \to \mathbb{R}^d \otimes \mathbb{R}^d$ be continuous. For $t \geq 0$ and $x \in E$, let $P_t(x, \cdot)$ be probability measures satisfying*

$$\begin{aligned}
&\int_E P_t(x, dy)(y_i - x_i) = b_i(x)\, t + o(t) \\
&\int_E P_t(x, dy)(y_i - x_i)(y_j - x_j) = a_{ij}(x)\, t + o(t) \qquad \text{(B.1.1)} \\
&\int_E P_t(x, dy) 1_{\{|x-y|>\varepsilon\}} = o(t),
\end{aligned}$$

uniformly in x as $t \to 0$ for all $i, j = 1, \ldots, d$ and $\varepsilon > 0$. Let A be the differential operator in (A.3.5). For $f \in \mathcal{C}(E)$ and $t > 0$, let $A_t f$ be given by

$$(A_t f)(x) := t^{-1}\Big(\int_E P_t(x, dy) f(y) - f(x)\Big). \qquad \text{(B.1.2)}$$

Then, for all $f \in \mathcal{C}^2(E)$,

$$\lim_{t\to 0} \sup_{x\in E} \Big|(A_t f)(x) - (Af)(x)\Big| = 0. \qquad \text{(B.1.3)}$$

Proof of Proposition B.1.1: Expand f in a Taylor series:

$$f(x + z) = f(x) + \sum_i z_i \tfrac{\partial}{\partial x_i} f(x) + \tfrac{1}{2} \sum_{ij} z_i z_j \tfrac{\partial^2}{\partial x_i \partial x_j} f(x) + R_x(z) \qquad \text{(B.1.4)}$$

and write the error terms in (B.1.1) as

$$\begin{aligned}\int_E P_t(x,dy)(y_i - x_i) &= b_i(x)\,t + o_i(x)(t)\\ \int_E P_t(x,dy)(y_i - x_i)(y_j - x_j) &= a_{ij}(x)\,t + o_{ij}(x)(t)\\ \int_E P_t(x,dy)1_{\{|x-y|>\varepsilon\}} &= o_\varepsilon(x)(t).\end{aligned}\tag{B.1.5}$$

Define

$$\begin{aligned} R_\varepsilon &:= \sup_{x\in E}\sup_{|z|\le\varepsilon} z^{-2}|R_x(z)|\\ T &:= \sup_{x,y\in E} |R_x(y-x)|\\ r'_t &:= t^{-1}\sup_{x\in E}\sqrt{\sum_i |o_i(x)(t)|^2}\\ r''_t &:= t^{-1}\sup_{x\in E}\sqrt{\sum_{ij} |o_{ij}(x)(t)|^2}\\ r_{\varepsilon,t} &:= t^{-1}\sup_{x\in E}|o_\varepsilon(x)(t)|\\ \|f'\| &:= \sup_{x\in E}\sqrt{\sum_i |\tfrac{\partial}{\partial x_i} f(x)|^2}\\ \|f''\| &:= \sup_{x\in E}\sqrt{\sum_{ij} |\tfrac{\partial^2}{\partial x_i\partial x_j} f(x)|^2}\\ \alpha &:= \sup_{x\in E}\sum_i a_{ii}(x).\end{aligned}\tag{B.1.6}$$

Then

$$\begin{aligned}\Big|(A_t f)(x) - (Af)(x)\Big| \le\; & r'_t\|f'\| + \tfrac{1}{2}r''_t\|f''\| + r_{\varepsilon,t}T\\ & + R_\varepsilon t^{-1}\int_E P_t(x,dy)|y-x|^2.\end{aligned}\tag{B.1.7}$$

so that for every $\varepsilon > 0$

$$\limsup_{t\to 0}\sup_{x\in E}\Big|(A_t f)(x) - (Af)(x)\Big| \le \alpha R_\varepsilon. \tag{B.1.8}$$

The result now follows from the fact that R_ε tends to zero as $\varepsilon \to 0$, i.e., for the error term in (B.1.4) one has

$$\lim_{\varepsilon\to 0}\sup_{x\in E}\sup_{|z|\le\varepsilon} z^{-2}|R_x(z)| = 0, \tag{B.1.9}$$

as follows from the uniform continuity of $\frac{\partial^2}{\partial x_i \partial x_j} f$ and the expression

$$R_x(z) = \sum_{ij} z_i z_j \int_0^1 d\lambda \int_0^\lambda d\mu \Big\{ \big(\tfrac{\partial^2}{\partial x_i \partial x_j} f\big)(x + \mu z) - \big(\tfrac{\partial^2}{\partial x_i \partial x_j} f\big)(x) \Big\}. \quad \text{(B.1.10)}$$

∎

B.2 Roots of diffusion matrices

Proposition B.2.1 *Let $E \subset \mathbb{R}^d$ be compact and convex, and let $a : E \to \mathbb{R}^d \otimes \mathbb{R}^d$ be a continuous function taking values in the $d \times d$ positive symmetric real matrices. Then there exists a continuous function $\sigma : E \to \mathbb{R}^d \otimes \mathbb{R}^d$ such that*

$$a_{ij}(x) = \sum_k \sigma_{ik}(x) \sigma_{jk}(x) \qquad (x \in E). \quad \text{(B.2.1)}$$

If, in addition, a is Lipschitz continuous and there exists an $\varepsilon > 0$ such that

$$\sum_{ij} z_i a_{ij}(x) z_j \geq \varepsilon |z|^2 \quad \forall z \in \mathbb{R}^d, \ \ x \in E, \quad \text{(B.2.2)}$$

then σ can be chosen Lipschitz continuous.

Here we equip the space $\mathbb{R}^d \otimes \mathbb{R}^d$ of $d \times d$ real matrices with the norm

$$\|a\|^2 := tr(a^\dagger a) = \sum_{ij} a_{ij} a_{ij}. \quad \text{(B.2.3)}$$

We say that a function $a : E \to \mathbb{R}^d \otimes \mathbb{R}^d$ is Lipschitz continuous if there exists a constant L such that

$$\|a(x) - a(y)\| \leq L|x - y| \qquad \forall x, y \in E. \quad \text{(B.2.4)}$$

Note that all norms on the finite-dimensional space $\mathbb{R}^d \otimes \mathbb{R}^d$ are equivalent, so that the definition of Lipschitz continuity does not depend on our choice for the norm $\|\cdot\|$ (only the Lipschitz constant does).

Proof of Proposition B.2.1: Any symmetric real matrix can be diagonalized. In particular, a positive symmetric a can be written as

$$a = \sum_i \lambda_i P_{\phi_i}, \quad \text{(B.2.5)}$$

where the ϕ_i form an orthonormal basis for $\mathbb{R}^d$, P_{ϕ_i} is the orthogonal projection on ϕ_i, and the λ_i are non-negative numbers. By definition the positive symmetric root $\sqrt{a}$ of a positive symmetric matrix a is

$$\sqrt{a} := \sum_i \sqrt{\lambda_i} P_{\phi_i}. \tag{B.2.6}$$

One can show that $\sigma := \sqrt{a}$ is the unique *positive symmetric* matrix such that

$$a_{ij} = \sum_k \sigma_{ik}\sigma_{jk}. \tag{B.2.7}$$

(There exist, however, other roots of a that are not positive symmetric.) We will show that the choice

$$\sigma(x) := \sqrt{a(x)} \tag{B.2.8}$$

satisfies the requirements in the Proposition.

To that end, we first note that whenever a and a' are symmetrix real matrices, and $a = \sum_i \lambda_i P_{\phi_i}$ and $a' = \sum_j \mu_j P_{\psi_j}$, then

$$\|a - a'\|^2 = \sum_{ij} |\lambda_i - \mu_j|^2 |\langle \phi_i | \psi_j \rangle|^2, \tag{B.2.9}$$

where $\langle \cdot | \cdot \rangle$ denotes the usual inner product on $\mathbb{R}^d$. With the help of this fact it is not hard to show that for any two positive symmetric matrices a and a'

$$\|\sqrt{a} - \sqrt{a'}\| \leq \sqrt{\|a - a'\|}. \tag{B.2.10}$$

If the spectrum of a and a' is bounded away from zero, in particular, if there exists an $\varepsilon > 0$ such that

$$\sum_{ij} z_i a_{ij} z_j \geq \varepsilon \sum_i z_i z_i \qquad \forall z \in \mathbb{R}^d, \tag{B.2.11}$$

and similarly for a', then it is even true that

$$\|\sqrt{a} - \sqrt{a'}\| \leq \frac{1}{2\varepsilon} \|a - a'\|. \tag{B.2.12}$$

Formula (B.2.10) proves that the function $x \mapsto \sigma(x)$ with σ as in (B.2.8) is continuous, and formula (B.2.12) gives us the Lipschitz continuity if a is Lipschitz continuous. ∎

Proposition B.2.2 *Let $E \subset \mathbb{R}^d$ be compact and convex, let $\sigma : E \to \mathbb{R}^d \otimes \mathbb{R}^d$ take values in the $d \times d$ real matrices, and assume that*

$$a_{ij}(x) = \sum_k \sigma_{ik}(x)\sigma_{jk}(x) \qquad (x \in E). \tag{B.2.13}$$

If there exists a direction $z \in \mathbb{R}^d$ with $|z| = 1$ and points $x_i, x \in E$ such that $x_i \to x$ as $i \to \infty$,

$$\sum_{ij} z_i a_{ij}(x) z_j = 0, \tag{B.2.14}$$

and

$$\limsup_{i\to\infty} |x_i - x|^{-2} \sum_{ij} z_i a_{ij}(x) z_j = \infty, \tag{B.2.15}$$

then $x \mapsto \sigma(x)$ is not Lipschitz continuous.

Proof of Proposition B.2.2: Assume that σ is Lipschitz continuous. Then the function $\psi : E \to \mathbb{R}^d$ given by

$$\psi_k(x) := \sum_i z_i \sigma_{ik}(x) \tag{B.2.16}$$

is Lipschitz continuous, and hence there exists a constant L such that

$$|\psi(x_i) - \psi(x)| \leq L|x_i - x| \qquad \forall i. \tag{B.2.17}$$

Here

$$|\psi(x)|^2 = \sum_{ijk} z_i \sigma_{ik}(x)\sigma_{jk}(x) z_j = 0, \tag{B.2.18}$$

and hence $\psi(x) = 0$. Inserting this into (B.2.17) we see that

$$\sum_{ij} z_i a_{ij}(x) z_j = |\psi(x_i)|^2 \leq L^2 |x_i - x|^2, \tag{B.2.19}$$

which contradicts (B.2.15). ∎

Bibliography

[1] J.-B. Baillon, Ph. Clément, A. Greven and F. den Hollander, On the attracting orbit of a non-linear transformation arising from renormalization of hierarchically interacting diffusions. Part I: The compact case. Can. J. Math. 47 (1995) 3–27.

[2] J.-B. Baillon, Ph. Clément, A. Greven and F. den Hollander, On the attracting orbit of a non-linear transformation arising from renormalization of hierarchically interacting diffusions. Part II: The non-compact case. J. Funct. Anal. 146 (1997) 236–298.

[3] L. Breiman, *Probability*, Addison-Wesley, Reading, 1968.

[4] K.L. Chung and R.J. Williams, *Introduction to Stochastic Integration*, Birkhäuser, Boston, 1990.

[5] J.T. Cox, K. Fleischmann and A. Greven, Comparison of interacting diffusions and applications to their ergodic theory. Probab. Theory Relat. Fields 105 (1996) 513–528.

[6] J.T. Cox and A. Greven, Ergodic theorems for infinite systems of locally interacting diffusions. Ann. Probab. 22 (1994) 833–853.

[7] J.T. Cox and D. Griffeath, Diffusive clustering in the two dimensional voter model. Ann. Probab. 14 (1986) 347–370.

[8] D.A. Dawson and A. Greven, Multiple time scale analysis of hierarchically interacting systems. In: *Stochastic Processes*, A Festschrift in Honour of Gopinath Kallianpur (eds. S. Cambanis, J.K. Ghosh, R.L. Karandikar and P.K. Sen), Springer, New York, 1993, pp. 41–50.

[9] D.A. Dawson and A. Greven, Multiple time scale analysis of interacting diffusions. Probab. Theory Relat. Fields 95 (1993) 467–508.

[10] D.A. Dawson and A. Greven, Hierarchical models of interacting diffusions: multiple time scale phenomena, phase transition and pattern of cluster-formation. Probab. Theory Relat. Fields 96 (1993) 435–473.

[11] D.A. Dawson and A. Greven, Multiple space-time scale analysis for interacting branching models. Electronic J. of Prob. 1 (1996), paper no. 14, pp. 1–84 (http://math.washington.edu/~ejpecp/).

[12] D.A. Dawson, A. Greven and J. Vaillancourt, Equilibria and quasi-equilibria for infinite collections of interacting Fleming-Viot processes. Trans. Am. Math. Soc. 347 (1995) 2277–2360.

[13] D.A. Dawson and P. March, Resolvent estimates for Fleming-Viot operators and uniqueness of solutions to related martingale problems. J. Funct. Anal. 132-II (1995) 417–472.

[14] D.A. Dawson and E.A. Perkins, Long-time behaviour and co-existence in a mutually catalytic branching model. Ann. Probab. 26 (1998) 1088–1138.

[15] E.B. Dynkin, *Markov Processes*, Die Grundlehren der mathematischen Wissenschaften 121–122, Springer, Berlin, 1965.

[16] S.N. Ethier and T.G. Kurtz, *Markov Processes; characterization and convergence*, John Wiley & Sons, New York, 1986.

[17] K. Fleischmann and A. Greven, Diffusive clustering in an infinite system of hierarchically interacting diffusions. Probab. Theory Relat. Fields 98 (1994) 517–566.

[18] K. Fleischmann and A. Greven, Time-space analysis of the cluster-formation in interacting diffusions. Electronic J. Probab. 1 (1996) 1–46.

[19] A. Friedman, *Partial Differential Equations of Parabolic Type*, Prentice-Hall, Englewood Cliffs, 1964.

[20] T. Hida, *Brownian motion*, Springer, New York, 1980.

[21] F. den Hollander and J.M. Swart, Renormalization of hierarchically interacting isotropic diffusions. J. Stat. Phys. 93 (1998) 243–291.

[22] I. Karatzas and S.E. Shreve, *Brownian Motion and Stochastic Calculus*, second edition, Springer, New York, 1991.

[23] A. Klenke, Different clustering regimes in systems of hierarchically interacting diffusions. Ann. Probab. 24 (1996) 660–697.

[24] U. Krengel, *Ergodic Theorems*, Walter de Gruyter, Berlin, 1985.

[25] T.M. Liggett, *Interacting Particle Systems*, Springer, New York, 1985.

[26] M. Notohara and T. Shiga, Convergence to genetically uniform state in stepping stone models of population genetics. J. Math. Biol. 10 (1980) 281–294.

[27] B. Oeksendal, *Stochastic Differential Equations: An Introduction with Applications*, Springer, Berlin, 1985.

[28] T. Ohta and M. Kimura, A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. Genet. Res. Camb. 22 (1973) 201–204.

[29] M.H. Protter and H.F. Weinberger, *Maximum Principles in Differential Equations*, Prentice-Hall, Englewood Cliffs, 1967.

[30] M. Reed and B. Simon, *Methods of Modern Mathematical Physics, I: Functional Analysis* (revised and enlarged edition), Academic Press, San Diego, 1980.

[31] D. Revuz and M. Yor, *Continuous Martingales and Brownian Motion*, Springer, Berlin, 1991.

[32] K.-I. Sato, Convergence to a diffusion of a multi-allelic model in population genetics. Adv. Appl. Prob. 10 (1978) 538–562.

[33] S. Sawyer and J. Felsenstein, Isolation by distance in a hierarchically clustered population. J. Appl. Probab. 20 (1983) 1–10.

[34] T. Shiga, An interacting system in population genetics. J. Math. Kyoto Univ. 20 (1980) 213–243.

[35] T. Shiga, An interacting system in population genetics II. J. Math. Kyoto Univ. 20 (1980) 723–733.

[36] T. Shiga, Diffusion processes in population genetics. J. Math. Kyoto Univ. 21 (1981) 133–151.

[37] T. Shiga, A certain class of infinite-dimensional diffusion processes arising in population genetics. J. Math. Soc. Japan 39 (1987) 17–25.

[38] T. Shiga, Ergodic theorems and exponential decay of sample paths for certain interacting diffusion systems. Osaka J. Math. 29 (1992) 789-807.

[39] T. Shiga and A. Shimizu, Infinite dimensional stochastic differential equations and their applications. J. Math. Kyoto Univ. 20 (1980) 395–416.

[40] D.W. Stroock and S.R.S. Varadhan, *Multidimensional Diffusion Processes*, Springer, Berlin, 1979.

[41] J.M. Swart, Clustering of linearly interacting diffusions and universality of their long-time distribution. Submitted to Prob. Th. Relat. Fields.

Samenvatting

Dit proefschrift, getiteld 'Het Gedrag van Lineair Wisselwerkende Diffusies op Grote Ruimte-Tijd-Schalen', is gewijd aan systemen die bestaan uit een aftelbaar oneindige collectie diffusies met een lineaire aantrekkende wisselwerking. Om precies te zijn beschouwen we oplossingen van stochastische differentiaalvergelijkingen van de volgende vorm:

$$dX_i(t) = \sum_j a(j-i)(X_j(t) - X_i(t))dt + \sigma(X_i(t))dB_i(t) \quad (i \in \Lambda).$$

Hierbij is Λ een of ander rooster (in het algemeen een Abelse groep) en hoort bij ieder roosterpunt $i \in \Lambda$ een diffusie $(X_i(t))_{t \geq 0}$. Deze diffusies zijn toevalsprocessen die worden aangestuurd door twee mechanismen. Ten eerste de ruistermen $\sigma(X_i(t))dB_i(t)$, die voor toevallige bewegingen zorgen die voor alle diffusies onafhankelijk zijn. Ten tweede de wisselwerkingstermen $\sum_j a(j-i)(X_j(t)-X_i(t))dt$, die elke diffusie in de richting van de waarden van de hem omringende diffusies drijven.[1]

Nadat in de inleiding (deel 1.1) is uitgelegd hoe zulke systemen ontstaan als de limieten van bepaalde discrete processen, worden in de hoofdstukken 2–4 enige aspecten van hun gedrag behandeld. Leitmotiv is daarbij het begrip universaliteit. Het blijkt dat systemen die verschillen in de ruis σ of de wisselwerking a, vaak toch ongeveer hetzelfde gedrag vertonen wanneer men ze beschouwt na lange tijd, of op grote ruimte-tijd-schalen. Deze bewering wordt precies gemaakt in verschillende limietstellingen.

Het gedrag na lange tijd wordt bijvoorbeeld beschreven in stelling 3.1.5, die zegt dat onder bepaalde voorwaarden de verdeling van het systeem als geheel $X(t) = (X_i(t))_{i \in \Lambda}$ convergeert naar een limietverdeling als $t \to \infty$. Het blijkt dat systemen met verschillende ruis en wisselwerking vaak dezelfde lange-tijd-limietverdeling hebben.

[1] Om precies te zijn: de diffusies $X_i(t)$ nemen waarden aan in een convexe verzameling $K \subset \mathbb{R}^d$, $\sigma : K \to \mathbb{R}^d \otimes \mathbb{R}^d$ is een matrixwaardige functie en $a : \Lambda \to [0, \infty)$ is de sprong-kern van een niet-reduceerbare recurrente toevalswandeling op Λ.

Het gedrag op grote ruimte-tijd-schalen wordt bijvoorbeeld beschreven in stelling 4.1.3. Daar beschouwen we 'blokgemiddelden' $\Lambda_n^{-1}\sum_{i\in\Lambda_n} X_i(\beta_n t)$, waarbij de $\Lambda_n \subset \Lambda$ grote blokken van roosterpunten zijn, en de β_n grote getallen die zorgen dat de tijd op de juiste manier wordt geschaald. We kunnen Λ_n en β_n op zo'n manier kiezen dat in de limiet $n \to \infty$ grote blokgemiddelden op een universele manier bewegen. In het bijzonder hangt de 'globale' ruis van een blokgemiddelde niet af van de 'locale' ruis van de individuele diffusies, binnen een geschikt gekozen universaliteitsklasse.

Het bewijzen van dergelijke stellingen over lineair wisselwerkende diffusies is niet nieuw. In 1980 behandelde Shiga [34] al het gedrag na lange tijd van diffusies $X_i(t)$ die waarden aannemen in $[0, 1]$ en waarvoor de ruisfunctie de vorm $\sigma(x) = \sqrt{x(1-x)}$ heeft, zogenaamde Wright-Fisher-diffusies. Dit is later uitgebreid naar algemenere diffusies op $[0, 1]$, waarbij een vergelijking met het Wright-Fisher-geval een belangrijke rol bleef spelen in het bewijs (zie [6]). Ook meerdimensionale en zelfs oneindig-dimensionale Wright-Fisher-diffusies werden behandeld in [12]. Het gedrag op grote ruimte-tijd-schalen voor diffusies op $[0, 1]$ werd behandeld met renormalisatie-achtige technieken in [1, 10]. Ook hier bleek het Wright-Fisher-geval weer een sleutelrol te spelen, dit keer als aantrekkend vast punt van een renormalisatietransformatie.

Nieuw in dit proefschrift is de behandeling van isotrope diffusies op algemene compacte en convexe domeinen $K \subset \mathbb{R}^d$. (Een isotrope diffusie op een niet-compact twee-dimensionaal domein werd gelijktijdig aan de productie van dit proefschrift behandeld in [14].) Voor deze diffusies kunnen we stellingen bewijzen over hun gedrag na lange tijd en op grote ruimte-tijd-schalen die analoog zijn aan de bekende resultaten voor de Wright-Fisher-diffusies. De gemeenschappelijke eigenschap van isotrope en Wright-Fisher-achtige diffusies die dat mogelijk maakt is een invariantie-eigenschap van hun harmonische functies. In het algemene, meerdimensionale geval is de ruisfunctie σ matrixwaardig. Zij $w := \sigma\sigma^T$, en zij A de differentiaaloperator

$$(Af)(x) := \sum_{\alpha\beta} w_{\alpha\beta}(x)\tfrac{\partial^2}{\partial x_\alpha \partial x_\beta} f(x) \quad (x \in K,\ f \in \mathcal{C}^2(K)).$$

De eigenschap die de isotrope diffusies en de Wright-Fisher-achtige diffusies gemeen hebben is dat de klasse van A-harmonische functies, d.w.z. de ruimte $H = \{f : Af = 0\}$ gesloten is onder transformaties van de vorm $f \mapsto \tilde{f}$, waarbij

$$\tilde{f}(x) := f(\lambda(x - y) + y),$$

met $\lambda \in [0, 1]$ en $y \in K$. Wanneer dit het geval is zeggen we dat de diffusies 'invariante harmonische functies' hebben.

Het blijkt nu dat we stellingen kunnen bewijzen over het gedrag na lange tijd en op grote ruimte-tijd-schalen voor systemen die invariante harmonische functies hebben. Dit gedrag blijkt bovendien universeel te zijn in de klasse van alle systemen die dezelfde klasse H van harmonische functies hebben. Zo is er een universaliteitsklasse van isotrope diffusies, en een van Wright-Fisher-achtige diffusies, en er zijn er nog meer.

Voor het gedrag na lange tijd is dit uitgewerkt in hoofdstuk 3. Tot nu toe was er niet veel aandacht voor de universaliteit van het gedrag na lange tijd, omdat dit voor de Wright-Fisher-achtige diffusies nogal triviaal is. Er blijkt echter een sterk verband te bestaan tussen de universaliteit na lange tijd en de universaliteit op grote ruimte-tijd-schalen. Formule (3.1.52) geeft aan hoe de universele ruis van grote blokgemiddelden kan worden uitgedrukt in de universele lange-tijd verdeling van het systeem. In hoofdstuk 4 (deel 4.4.1) wordt deze formule verklaard.

De belangrijkste conclusies van dit proefschrift zijn dat universeel gedrag van lineair wisselwerkende diffusies, na lange tijd en op grote ruimte-tijd-schalen, optreedt als de klasse H van harmonische functies gesloten is onder transformaties zoals hierboven beschreven, en dat het de moeite waard is om stil te staan bij de universaliteit van het gedrag na lange tijd, omdat dit nauw samenhangt met de (moeilijkere en diepere) universaliteit op grote ruimte-tijd-schalen.

Summary

This dissertation, with the title 'Large Space-Time Scale Behavior of Linearly Interacting Diffusions', is devoted to systems consisting of a countably infinite collection of diffusions with a linear attractive interaction. More precisely we consider solutions of stochastic differential equations of the following form:

$$dX_i(t) = \sum_j a(j-i)(X_j(t) - X_i(t))dt + \sigma(X_i(t))dB_i(t) \quad (i \in \Lambda).$$

Here Λ is some lattice (in the general case an Abelian group) and with each point $i \in \Lambda$ there is associated a diffusion $(X_i(t))_{t\geq 0}$. These diffusions are stochastic processes that are driven by two mechanisms. Firstly, the noise terms $\sigma(X_i(t))dB_i(t)$, which cause random movements that are independent for all diffusions. Secondly, the interaction terms $\sum_j a(j-i)(X_j(t)-X_i(t))dt$, which drive each diffusion in the direction of the values of the diffusions that surround it.[2]

After the Introduction, where (in section 1.1) it is explained how such systems arise as the limits of certain discrete processes, in the Chapters 2–4 some aspects of their behavior is treated. The guiding theme is here the concept of universality. It turns out that systems that differ in the noise σ or the interaction a, often show approximately the same behavior when viewed after a long time, or on large space-time scales. This statement is made precise in several limit theorems.

The long-time behavior is described, for example, in Theorem 3.1.5, which says that under certain assumptions the law of the system as a whole $X(t) = (X_i(t))_{i\in\Lambda}$ converges to a limiting law as $t \to \infty$. It turns out that systems that differ in their noise and interaction often share the same long-time limiting law.

The behavior on large space-time scales is described, for example, in Theorem 4.1.3. There we consider 'block averages' $\Lambda_n^{-1}\sum_{i\in\Lambda_n} X_i(\beta_n t)$, where the $\Lambda_n \subset \Lambda$ are large blocks containing points in the lattice, and the β_n are large numbers that rescale time in the correct way. We can choose Λ_n and β_n in such a way

[2]To be precise: the diffusions $X_i(t)$ take values in a compact set $K \subset \mathbb{R}^d$, $\sigma : K \to \mathbb{R}^d \otimes \mathbb{R}^d$ is a matrix-valued function and $a : \Lambda \to [0, \infty)$ is the jump kernel of an irreducible recurrent random walk on Λ.

that, in the limit $n \to \infty$, large block averages move in a universal way. In particular we show that the 'global' noise of the block average does not depend on the 'local' noise of the individual diffusions, for systems in an appropriate universality class.

To prove this type of theorems on linearly interacting diffusions is not new. In 1980 Shiga [34] already treated the long-time behavior of diffusions $X_i(t)$ taking values in $[0, 1]$ and subject to a noise function of the form $\sigma(x) = \sqrt{x(1-x)}$, so-called Wright-Fisher diffusions. This was later extended to more general diffusions on $[0, 1]$, where a comparison to the Wright-Fisher case still played an important role in the proof (see [6]). Also more-dimensional and even infinite-dimensional Wright-Fisher diffusions were treated in [12]. The behavior on large space-time scales for diffusions on $[0, 1]$ was treated with renormalization techniques in [1, 10]. Here too the Wright-Fisher case played a key role, this time as the attractive fixed point of a renormalization transformation.

New in this dissertation is the treatment of isotropic diffusions on general compact and convex domains $K \subset \mathbb{R}^d$. (An isotropic diffusion on a non-compact two-dimensional domain was treated [14] at the same time as this dissertation was produced.) For these diffusions we can prove theorems on their long-time and large space-time scale behavior that are analogous to the known results for Wright-Fisher diffusions. The properties that isotropic and Wright-Fisher diffusions have in common and that allows this is an invariance property of their harmonic functions. In the general, more-dimensional case the noise function σ is matrix-valued. Let $w := \sigma\sigma^T$, and let A be the differential operator

$$(Af)(x) := \sum_{\alpha\beta} w_{\alpha\beta}(x) \tfrac{\partial^2}{\partial x_\alpha \partial x_\beta} f(x) \quad (x \in K,\ f \in \mathcal{C}^2(K)).$$

The property shared by isotropic diffusions and Wright-Fisher diffusions is the fact that the class of A-harmonic functions, i.e., the space $H = \{f : Af = 0\}$, is closed under transformations of the form $f \mapsto \tilde{f}$, where

$$\tilde{f}(x) := f(\lambda(x-y)+y),$$

with $\lambda \in [0, 1]$ and $y \in K$. If this is the case, then we say that the diffusions have 'invariant harmonics'.

It now turns out that we can prove theorems on the long-time and large space-time scale behavior of systems with invariant harmonics. Moreover, this behavior turns out to be universal in the class of all diffusions that share the same class H of harmonic functions. In this way there is a universality class of Wright-Fisher-type diffusions, one of isotropic diffusions, and there are more of them.

For the long-time behavior this idea is developed in Chapter 3. Till recently the universality of long-time behavior received little attention, because of the fact that it is rather trivial for Wright-Fisher-type diffusions. It turns out, however, that there exists a strong link between long-time universality and large space-time scale universality. Formula (3.1.52) expresses how the universal noise of large block averages can be expressed in the universal long-time distribution of the system. In Chapter 4 (section 4.4.1) this formula is explained.

The most important conclusions of this dissertation are that universal behavior of linearly interacting diffusions, both after long time and on large space-time scales, occurs if the class H of harmonic functions is closed under transformations as described above, and that it pays off to give some attention to the universal nature of long-time behavior, because it is closely linked to the (more difficult and deeper) universality on large space-time scales.

Acknowledgements

I would like to thank all those people with whom I discussed maths during the last four years, and who contributed to this dissertation, either because they commented on parts of the text, or simply because they answered questions or posed their own ones. I think of Philippe Clément, who asked ‘what are your boundary conditions?’, when it had never occured to me how remarkable it was that we did not need any. Often it is so much easier to learn in interaction with an expert, than from a silent book or article. I remember how easy regularity theory for solutions of Cauchy equations sounded when Tom ter Elst explained it, and how he gave me some articles that I am still unable to read. I remember Achim Klenke, who in one email admitted that the ‘immediate consequence’ I had been puzzled by for a day was in fact ‘sort of immediate’, if only you had the right idea, and then went on to give a nice and elegant proof. And I remember E. Perkins, who helped us in a crucial way when he wrote that the uniqueness problem we had been struggling with for so long ‘is a well-known problem in the field, and many experts have been working on it’. In the same way I must thank S.N. Ethier and T. Shiga, and many others. Here in Nijmegen I thank my part-time collegue and companion Marek Biskup; furthermore R. Kortram, A. van Rooij and W. Pestman, who always welcomed someone walking into their office with a question. Wiebe’s phenomenal knowledge about measure theory and functional analysis will be missed in Nijmegen.

I owe special thanks to Andreas Greven, with whom I could discuss every aspect of the linearly interacting diffusions, from the vaguest intuition to the concrete proofs. There are also special thanks to Frank den Hollander, for putting up with me for four years, for reading carefully those hundreds of pages of previous versions and notes that are not part of this dissertation now, and for teaching me also all those non-mathematical skills that a mathematician needs.

Finally, I would like to thank all friends who did not run away during the last years, in spite of a gross lack of attention.

Index

Curriculum Vitae

I was born on June 7th 1970 in Drachten, The Netherlands, where I attended the gymnasium. After perticipating in the International Physics Olympiade in the Summer of 1988, I went to Groningen to study physics. In 1994 I graduated cum laude in theoretical physics. My master's thesis tried to give a definition of information transfer in terms of conditional independence relations, and was linked to the philosophy and interpretation of quantum mechanics. It also led (in 1995) to a four-page paper in Statistics & Probability Letters, on the (non-) existence of conditional product measures. After some time of fruitlessly looking for a PhD position in the foundations and philosophy of physics, I turned to the 'honest craftmanship' of mathematics in 1995, when I started to work as an OiO on the four-year project 'Renormalization of Interacting Diffusions', the result of which is the present dissertation. During those four years I had regular contacts with Andreas Greven and his group in Erlangen, Germany. In the late Summer of 1998 I spent two months at the Fields Institute in Toronto. I gave talks in Toronto, Utrecht, Amsterdam and Berlin and visited several seminars and workshops. Since October 1st, 1999 I am working as a postdoc at the TU Berlin, in the Graduiertenkolleg 'Stochastische Prozesse und probabilistische Analysis' where I am happy to find several people eager to attack with me some of the open problems in the linear interacting diffusions business.